

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:30:02 ; Search time 6.00576 Seconds
(without alignments)
2914.068 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_340

Perfect score: 1744

Sequence: 1 AEEGAVAVCVVRPLNSREE.....QFASTAKYMKNTPYVNEVST 339

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgm2_6/ptodata/2/iaa/5A_COMB.pap.*
- 2: /cgm2_6/ptodata/2/iaa/5B_COMB.pap.*
- 3: /cgm2_6/ptodata/2/iaa/6A_COMB.pap.*
- 4: /cgm2_6/ptodata/2/iaa/6B_COMB.pap.*
- 5: /cgm2_6/ptodata/2/iaa/PTUS_COMB.pap.*
- 6: /cgm2_6/ptodata/2/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1744	100.0	2662	4	US-09-595-684B-31
2	1299.5	74.5	2954	4	US-09-150-867-1
3	668	38.3	473	4	US-09-592-054-6
4	668	38.3	522	4	US-09-592-054-4
5	658	37.7	1231	4	US-09-595-684B-23
6	652	37.4	1232	4	US-09-592-054-2
7	651	37.3	1234	4	US-09-592-054-8
8	615.5	35.3	411	2	US-08-713-815A-4
9	615.5	35.3	441	2	US-08-713-815A-3
10	615.5	35.3	975	4	US-09-914-259-19
11	615	35.3	409	4	US-09-592-191-6
12	615	35.3	409	4	US-09-723-262-6
13	615	35.3	409	4	US-09-723-219-6
14	615	35.3	1388	4	US-09-592-191-2
15	615	35.3	1388	4	US-09-723-262-2
16	615	35.3	1388	4	US-09-723-219-2
17	612.5	35.1	1031	4	US-09-914-259-24
18	605	34.7	1066	3	US-09-541-782-8
19	605	34.7	1066	4	US-09-723-820-8
20	605	34.7	1066	4	US-10-270-085-8
21	604	34.6	963	4	US-09-914-259-22
22	603	34.6	375	4	US-09-592-191-4
23	603	34.6	375	4	US-09-723-262-4
24	603	34.6	375	4	US-09-723-219-4
25	603	34.6	963	4	US-09-914-259-20
26	602.5	34.5	1032	4	US-09-914-259-26
27	600.5	34.4	1027	4	US-09-914-259-27

28	599	34.3	957	4	US-09-914-259-16	Sequence 16, Appl
29	598.5	34.3	967	4	US-09-914-259-21	Sequence 21, Appl
30	598	34.3	928	4	US-09-914-259-23	Sequence 23, Appl
31	595	34.1	956	4	US-09-914-259-17	Sequence 17, Appl
32	594.5	34.1	1279	4	US-09-724-517-2	Sequence 2, Appli
33	594.5	34.1	1279	4	US-09-641-807A-2	Sequence 2, Appli
34	594.5	34.1	1279	4	US-09-723-096-2	Sequence 2, Appli
35	591	33.9	815	4	US-09-914-259-18	Sequence 18, Appl
36	587.5	33.7	935	4	US-09-914-259-25	Sequence 25, Appl
37	581	33.3	1690	4	US-09-595-684B-39	Sequence 39, Appl
38	574.5	32.9	341	4	US-09-724-517-4	Sequence 4, Appli
39	574.5	32.9	341	4	US-09-641-807A-4	Sequence 4, Appli
40	574.5	32.9	341	4	US-09-723-096-4	Sequence 2, Appli
41	569.5	32.7	342	4	US-09-641-806-2	Sequence 2, Appli
42	569.5	32.7	342	4	US-09-723-129-2	Sequence 2, Appli
43	569.5	32.7	342	4	US-09-722-862-2	Sequence 2, Appli
44	569.5	32.7	1375	4	US-09-722-139-2	Sequence 2, Appli
45	569.5	32.7	1375	4	US-09-721-832-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; FILE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: PRT
; ORGANISM: Human
US-09-595-684B-31

Query Match 100.0%; Score 1744; DB 4; Length 2662;
Best Local Similarity 100.0%; Pred. No. 1.9e-176;
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AEEGAVAVCVVRPLNSREESLGETAQVYKTDNNVYQVDSKSFNFDFVHGNETKN 60
Db 2 AEEGAVAVCVVRPLNSREESLGETAQVYKTDNNVYQVDSKSFNFDFVHGNETKN 61
QY 61 VYBEIAAPIIDSAIOQYNGTIFAYGTAGCTYTMGSEDLGVIPRAIHDFQIKKFP 120
Db 62 VYBEIAAPIIDSAIOQYNGTIFAYGTAGCTYTMGSEDLGVIPRAIHDFQIKKFP 121
QY 121 DREFLARVSMETLYNETITDLCGTQMKPLIREDVNRNRYADLTTEVVYVSEMAK 180
Db 122 DREFLARVSMETLYNETITDLCGTQMKPLIREDVNRNRYADLTTEVVYVSEMAK 181
QY 181 ITKGEKSRHYGETKMQRSRSHITFRMILESEKGEPSNCEGSKVSHLNLVDLAGSER 240
Db 182 ITKGEKSRHYGETKMQRSRSHITFRMILESEKGEPSNCEGSKVSHLNLVDLAGSER 241
QY 241 AAQTGAAGVRLKEGCNINSLFLQGVIKLSDGVGGFYNFVYDSKLTILQNSLGNPK 300
Db 242 AAQTGAAGVRLKEGCNINSLFLQGVIKLSDGVGGFYNFVYDSKLTILQNSLGNPK 301

```

QY      301 TRIICITTPVSFDETLTALQFASTAKYMKNTPYVNEVST 339
       |||||||
Db      302 TRIICITTPVSFDETLTALQFASTAKYMKNTPYVNEVST 340
       |||||||

RESULT 2
US-09-150-867-1
; Sequence 1, Application US/09150867
; Patent No. 6645748
; GENERAL INFORMATION:
; APPLICANT: Wood, Kenneth W.
; APPLICANT: Sakowicz, Roman
; APPLICANT: Goldstein, Lawrence S.B.
; APPLICANT: Cleveland, Don W.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Plus End-Directed Microtubule Motor Required for
; FILE REFERENCE: 18557C-000110US
; CURRENT APPLICATION NUMBER: US/09/150,867
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: US 60/058,645
; EARLIER FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2954
; TYPE: PRT
; ORGANISM: Xenopus sp.
; FEATURE:
; OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)
; OTHER INFORMATION: member of the kinesin superfamily of microtubule
; OTHER INFORMATION: motor proteins
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1)..(472)
; OTHER INFORMATION: kinesin like motor domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (473)..(2752)
; OTHER INFORMATION: rod domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (2753)..(2954)
; OTHER INFORMATION: tail domain
US-09-150-867-1

Query Match          74.5%; Score 1299.5; DB 4; Length 2954;
Best Local Similarity 73.5%; Pred. No. 7,3e-129;
Matches 250; Conservative 42; Mismatches 43; Indels 5; Gaps 2;

QY      1 AREGAVACVRVRPLNSREESLGETAQVYWKTDNNVIYQVGSKSFNDRVPFHGNETTKN 60
       :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      2 SEGDAVKVCVRVRPLIQREQ--GDQANLOWKAGNNTISQVDGTKSFNDRVFNSESHTSQ 59
       :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      61 VTEETAAPIDSAIQGYNGTIFAYGOTASGKTYTWMGSEDLHGVIPRALHDIFQIKKFP 120
       :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      60 IYQEATAVPIRSALQGYNGTIFAYGQTSGGKYTWMTGPSNLUGIIPQAIQEVFKLIQIP 119
       :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      121 DREFILRVSYMIYNFTIIDLLCGTGQMKPLIREDVNRNVYADLTBEVVYTSMAALKW 180
       :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      120 NREFILRVSYMEIYNFTKVLDLCDDRRKKPLEIREDFNRNVYADLTTELVMVPEHVIQW 179
       :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      181 IYKGESRHYGETXONQRSRHPTFRMILESEKEEPS--NCEGSVKVSHLNLDVLG 237
       :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      180 IKKGKNRAYGETKMWDHSSRHPTFRMTVESGRDNDPTNSENDGVAMVSHLNLDVLG 239
       :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      238 SERAAQTGAAGVRLKEGCININSFLFGQVKKLSDGQVGCGINTYRDYSKLTRILQNSLGG 297
       :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      240 SERASTGTGAAGVRLKEGCININSLSFILGQVKKLSDGQAGGFINYRDSKLTRILQNSLGG 299
       :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      298 NPKTRICTITPPVSFDETLTALQFASTAKYMKNTPYVNEV 337
       :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      300 NAKTVIICTITPPVSFDETLSTLQFASTAKHVRNTPHVNVEV 339
       :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

US-09-592-054-4

Query Match 38.3%; Score 668; DB 4; Length 522;
Best Local Similarity 44.7%; Pred. No. 1.5e-62;
Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 5;

QY 6 VAVCVVRPLNRSRLG-ETAQVYWKTDNNVIYQVDGSKSFNDRVPHGNETTKNVEE 64
DB 30 VRVALRCRPLVPEISEGCMCLSFVGPQPQVVVGTG--KSFTYDFVDPSTQEVEFNT 87
QY 65 IAAPLIDSAIQYNGTIFAYGOTASGKTYTMG-----SEDHLGVIPRAIHDFQKIK 117
DB 88 AVAPLKGKGVFNATVLAAYGQTSKTYSMGAYTAQENEPVGVIPRVQLLFKEID 147
QY 118 KFPDRFLLRVSYMEIYNETITDLCGTOKMKPLIREDVNRNVVADLTVEVVYVTSMA 177
DB 148 KKSDFEFTLKVSYLEIYNEEILDLCPSEKQAQINREDPKGKIVGLTEKTVLVALDT 207
QY 178 LKWITGKSRHYGETKQNRSSRHTIFRMILESEKGEPCNSCEGSKVSHNLVLDLAG 237
DB 208 VSCLEQGNNSRTVASTAMNSQSSRSHAIPTISLEQKKSD----KNSSFRSKLHLVLDLAG 263
QY 238 SERAAQTGAAGVRLKEGNCINRSFLIGQVVKLSDGQVGVGFYINRDSKLTIRLQNSLGG 297
DB 264 SERQKTKAEGDRLKEGININRGLLCLGNVISALGDDKGGFVYRDSKLTIRLQNSLGG 323
QY 298 NPKTRIICTITPV--SFDETLTALQFASHTAKYMKNTPVYN 335
DB 324 NSHTLMIACVSPADSNLEETLTLRYADRAKIKNKPIVN 363

RESULT 5

US-09-595-684B-23
; Sequence 23, Application US/09595684B
; Patent No. 6544766

; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 1231
; TYPE: PRT
; ORGANISM: Human

US-09-595-684B-23

Query Match 37.7%; Score 658; DB 4; Length 1231;
Best Local Similarity 44.7%; Pred. No. 7e-61;
Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 6;

QY 6 VAVCVVRPLNRSRLG-ETAQVYWKTDNNVIYQVDGSKSFNDRVPHGNETTKNVEE 64
DB 10 VRVALRCRPLVPEISEGCMCLSFVGPQPQVVVGTG--KSFTYDFVDPSTQEVEFNT 67
QY 65 IAAPLIDSAIQYNGTIFAYGOTASGKTYTMG-----SEDHLGVIPRAIHDFQKIK 117
DB 68 AVAPLKGKGVFNATVLAAYGQTSKTYSMGAYTAQENEPVGVIPRVQLLFKEID 127
QY 118 KFPDRFLLRVSYMEIYNETITDLCGTOKMKPLIREDVNRNVVADLTVEVVYVTSMA 177
DB 128 KKSDFEFTLKVSYLEIYNEEILDLCPSEKQAQINREDPKGKIVGLTEKTVLVALDT 187

QY 178 LKWITGKSRHYGETKQNRSSRHTIFRMILESEKGEPCNSCEGSKVSHNLVLDLAG 237
DB 188 VSCLEQGNNSRTVASTAMNSQSSRSHAIPTISLEQKKSDKNSSFR-SKLHLVLDLAG 243
QY 238 SERAAQTGAAGVRLKEGNCINRSFLIGQVVKLSDGQVGVGFYINRDSKLTIRLQNSLGG 297
DB 244 SERQKTKAEGDRLKEGININRGLLCLGNVISALGDDKGGFVYRDSKLTIRLQNSLGG 303
QY 298 NPKTRIICTITPV--SFDETLTALQFASHTAKYMKNTPVYN 335
DB 304 NSHTLMIACVSPADSNLEETLTLRYADRAKIKNKPIVN 343

RESULT 6

US-09-592-054-2
; Sequence 2, Application US/09592054
; Patent No. 6440684

; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Finer, Jeffrey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6440684el motor proteins and methods for
; FILE REFERENCE: 1016
; CURRENT APPLICATION NUMBER: US/09/592,054
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1232
; TYPE: PRT
; ORGANISM: Human

US-09-592-054-2

Query Match 37.4%; Score 652; DB 4; Length 1232;
Best Local Similarity 44.4%; Pred. No. 3.1e-60;
Matches 151; Conservative 52; Mismatches 121; Indels 16; Gaps 6;

QY 6 VAVCVVRPLNRSRLG-ETAQVYWKTDNNVIYQVDGSKSFNDRVPHGNETTKNVEE 64
DB 10 VRVALRCRPLVPEISEGCMCLSFVGPQPQVVVGTG--KSFTYDFVDPSTQEVEFNT 67
QY 65 IAAPLIDSAIQYNGTIFAYGOTASGKTYTMG-----SEDHLGVIPRAIHDFQKIK 117
DB 68 AVAPLKGKGVFNATVLAAYGQTSKTYSMGAYTAQENEPVGVIPRVQLLFKEID 127
QY 118 KFPDRFLLRVSYMEIYNETITDLCGTOKMKPLIREDVNRNVVADLTVEVVYVTSMA 177
DB 128 KKSDFEFTLKVSYLEIYNEEILDLCPSEKQAQINREDPKGKIVGLTEKTVLVALDT 187
QY 178 LKWITGKSRHYGETKQNRSSRHTIFRMILESEKGEPCNSCEGSKVSHNLVLDLAG 237
DB 188 VSCLEQGNNSRTVASTAMNSQSSRSHAIPTISLEQKKSDKNSSFR-SKLHLVLDLAG 243
QY 238 SERAAQTGAAGVRLKEGNCINRSFLIGQVVKLSDGQVGVGFYINRDSKLTIRLQNSLGG 297
DB 244 SERQKTKAEGDRLKEGININRGLLCLGNVISALGDDKGGFVYRDSKLTIRLQNSLGG 303
QY 298 NPKTRIICTITPV--SFDETLTALQFASHTAKYMKNTPVYN 335
DB 304 NSHTLMIACVSPADSNLEETLTLRYADRAKIKNKPIVN 343

RESULT 7

US-09-592-054-8
; Sequence 8, Application US/09592054
; Patent No. 6440684

; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Finer, Jeffrey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth

;; TITLE OF INVENTION: No. 6440684e1 motor proteins and methods for

;; FILE REFERENCE: 1016

;; CURRENT APPLICATION NUMBER: US/09/592,054

;; CURRENT FILING DATE: 2000-07-20

;; NUMBER OF SEQ ID NOS: 8

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 8

;; LENGTH: 1234

;; TYPE: PRT

;; ORGANISM: Human

US-09-592-054-8

Query Match 37.3%; Score 651; DB 4; Length 1234;

Best Local Similarity 43.8%; Pred. No. 3.9e-60;

Matches 149; Conservative 56; Mismatches 119; Indels 16; Gaps 6;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDFRVHGNETTKVYEE 64

DB 10 VRVALRCRPLVPKEISEGCMCLSFVPGETQVVVGTD--KSFTYDFVDFPCTEQEEVFNK 67

QY 65 IAAPIIDSALQYNGTIFAYGQTASGKTYTMG-----SEDLHGVIPRAIHDFQKIK 117

DB 68 AVAPLIKGFPGYNATVLAAYGQTSGKTSYMGGAATAEQENETVGLIPRVIQLLPKEID 127

QY 118 KPDRFLLRVSYMEIYNETITDLGCTQKMKPLIREDVNRNVIYVADLTEEVEVYVYSEMA 177

DB 128 QKSDFTFLKVSLEIYNEELDLCPREKAQINIREDPKEGKIIVGLTEKTVLVALDT 187

QY 178 LKWTITGKSRHYGETKMNORSRSHTIPRMILESRKGPSPCEGSKVSHNLVDLAG 237

DB 188 VSCLEQNNRSRTVASTAMNSQSSRSHAFTI---SLEQKKSKDNSSFR-SKLHLVDLAG 243

QY 238 SERAAQTGAAGVRLKEGCNINRSFILGQVTKKLSGQVGFNYRDSKLTIRLQNSLGG 297

DB 244 SERQKTKARGDLKEGINNRGLCLGNVISALGDDKGSFVYRDSKLTIRLQNSLGG 303

QY 298 NPKTRIICTTPV--SPDELTALQFASTAKYMKNTPYVN 335

DB 304 NSHTLMIACVSPADSNLEETLSTLYADRAKIKNKIIVN 343

RESULT 8

US-08-713-815A-4

;; Sequence 4, Application US/08713815A

;; Patent No. 5830659

;; GENERAL INFORMATION:

;; APPLICANT: Russell J. Stewart

;; TITLE OF INVENTION: ACTIVE MICROTUBULE-BASED

;; TITLE OF INVENTION: SEPARATIONS BY KINESINS

;; NUMBER OF SEQUENCES: 6

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Thorpe, No. 5830659th & Western, L.L.P.

;; STREET: 9035 South 700 East, Suite 200

;; CITY: Sandy

;; STATE: Utah

;; COUNTRY: USA

;; ZIP: 84070

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage

;; COMPUTER: AST Ascentia 900N

;; OPERATING SYSTEM: DOS 6.22

;; SOFTWARE: Word Perfect 6.0

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/713,815A

;; FILING DATE: 13-SEP-1996

;; CLASSIFICATION: 435

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Alan J. Howarth

;; REGISTRATION NUMBER: 36,553

;; REFERENCE/DOCKET NUMBER: T3214/U-2202

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (801)566-6633

;; TELEFAX: (801)566-0750

;; INFORMATION FOR SEQ ID NO: 4:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 411 amino acid residues

;; TYPE: amino acid

;; TOPOLOGY: linear

US-08-713-815A-4

Query Match

Best Local Similarity 35.3%; Score 615.5; DB 2; Length 411;

Matches 158; Conservative 43; Mismatches 125; Indels 23; Gaps 12;

QY 1 AREGAVAVCVVRPLNSREESLG-ETAQVYWKTDNNV---IYQVDSKSFNDFRVHGNET 57

DB 9 AEDSIKVCV-RRPLNDSEKAGSKFVV--KFPNVVEENCISLAG-KVYLFQKVPKNAS 64

QY 58 TKNVVEEIAAPIDSALQYNGTIFAYGQTASGKTYTM---MGSEDLHGVIPRAIHDFQ 114

DB 65 QEKVYNEAKSIVTDVLAAGYNGTIFAYGQTSSGKTHMEGVIGDSVKQGIIPRVNDIFN 124

QY 115 KIKKFP-DREFLLRVSYMEIYNETITDLGCTQKMKPLIREDVNRNVIYVADLTEEVEVY 173

DB 125 HIYAMEVNLFEHIKVSYYEIMDKIRDLL-DVSKVN-LSVHEDKNRPVYKGAFTERVSS 182

QY 174 SEMALKWITGKSRHYGETKMNORSRSHTIPRMILESRKGPSPCEGSKVSVKVS-HLNL 232

DB 183 PEDVFEVIEGKSNRHIAVNNHESRSHSVFLINVKQ-----ENLENQKKLSGKLYL 236

QY 233 VDLAGSERAQGTGAAGVRLKEGCNINRSFILGQVTKKLSGQVGFNYRDSKLTIRLQ 292

DB 237 VDLAGSEKVSCTGAEGTVLDEAKNINKLSALGNVISALADGN-KTHIPYRDSKLTIRLQ 295

QY 293 NSLGGNPKTRIICTTPVSPD--ETLTALQFASTAKYMKNTPYVNEVST 339

DB 296 ESLGNGARTTIVCCSPASFNESETKSLDGFRRAKTVRNVCVNEELT 344

RESULT 9

US-08-713-815A-3

;; Sequence 3, Application US/08713815A

;; Patent No. 5830659

;; GENERAL INFORMATION:

;; APPLICANT: Russell J. Stewart

;; TITLE OF INVENTION: ACTIVE MICROTUBULE-BASED

;; TITLE OF INVENTION: SEPARATIONS BY KINESINS

;; NUMBER OF SEQUENCES: 6

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Thorpe, No. 5830659th & Western, L.L.P.

;; STREET: 9035 South 700 East, Suite 200

;; CITY: Sandy

;; STATE: Utah

;; COUNTRY: USA

;; ZIP: 84070

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage

;; COMPUTER: AST Ascentia 900N

;; OPERATING SYSTEM: DOS 6.22

;; SOFTWARE: Word Perfect 6.0

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/713,815A

;; FILING DATE: 13-SEP-1996

;; CLASSIFICATION: 435

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Alan J. Howarth

;; REGISTRATION NUMBER: 36,553

;; REFERENCE/DOCKET NUMBER: T3214/U-2202

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (801)566-6633

;; TELEFAX: (801)566-0750

;; INFORMATION FOR SEQ ID NO: 3:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 441 amino acid residues

;; TYPE: amino acid


```
; CURRENT APPLICATION NUMBER: US/09/723,262
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Human
; US-09-723-262-6

Query Match          35.3%; Score 615; DB 4; Length 409;
Best Local Similarity 43.3%; Pred. No. 4.8e-57;
Matches 158; Conservative 48; Mismatches 113; Indels 46; Gaps 10;

QY      2 EGAIVAVCVVRPLNSREESL-GETAQVYWKTDNNVIYQVDGS-----KSFNFD 49
Db      23 EGDALKVVRIRPPAERSGSADGE-----QNCLSVLSSTSLRLHSNPEPKTFTFD 73

QY      50 RVFHGNETTKNVEEIAAIIIDSAIQYNGTIFAYGQTASGKTYTMMG-----SEDHL 102
Db      74 HVADVDVTTQESVFATVAKSIVESCMGNGYNGTIFAYGQTGSGKFTTMMGPSEDNFSHNL 133

QY      103 GVIPRAIHDF-----QKIKFPDPREFLLRVSYMEIYNETITDLCGTQKMKPLIREDV 157
Db      134 GVIPRSFEYLFSLIDREKEKAGAKSFLCKCSFIEIYNEQIYDLL--DSASAGLYLREHI 191

QY      158 NRVVYVADLTEEVVYVTSSEMALKWITKGEKSRHVGKTMNQRSSRSHTIFFRMILESREKGE 217
Db      192 KKGVFVVGAVEQVVTSAEAYQVLGGWNRNRVASTSMNRESRSSHAVFTTIESMEK-- 249

QY      218 PSNCEGSKVSHNLNLDLAGSRAAQTGAAGVRLKEGCNINRSLFILGOVVKLSGQVVG 277
Db      250 -SNEIVNIRTSLLNLDLAGSERQKDTAEGMKLEAGNINRSLSCLGQVITALVD--VG 306

QY      278 G-----FINVRDSKLTILONSLGNGPKTRIICTITPVS--FDETLTALQFASAKYMKNT 331
Db      307 NGQRHVCYRDSKLTFLRDSLGGNAKTAIIANVHPGSRFCFGETLSTLNPAQAKLIKKN 366

QY      332 PYNE 336
Db      367 AVNE 371

RESULT 14
US-09-572-191-2
; Sequence 2, Application US/09572191
; Patent No. 6355466
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6355466el motor proteins and methods for
; FILE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/572,191
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
; US-09-572-191-2

Query Match          35.3%; Score 615; DB 4; Length 1388;
Best Local Similarity 43.3%; Pred. No. 3.4e-56;
Matches 158; Conservative 48; Mismatches 113; Indels 46; Gaps 10;

QY      2 EGAIVAVCVVRPLNSREESL-GETAQVYWKTDNNVIYQVDGS-----KSFNFD 49
Db      23 EGDALKVVRIRPPAERSGSADGE-----QNCLSVLSSTSLRLHSNPEPKTFTFD 73

QY      50 RVFHGNETTKNVEEIAAIIIDSAIQYNGTIFAYGQTASGKTYTMMG-----SEDHL 102
Db      74 HVADVDVTTQESVFATVAKSIVESCMGNGYNGTIFAYGQTGSGKFTTMMGPSEDNFSHNL 133

QY      103 GVIPRAIHDF-----QKIKFPDPREFLLRVSYMEIYNETITDLCGTQKMKPLIREDV 157
Db      134 GVIPRSFEYLFSLIDREKEKAGAKSFLCKCSFIEIYNEQIYDLL--DSASAGLYLREHI 191

QY      158 NRVVYVADLTEEVVYVTSSEMALKWITKGEKSRHVGKTMNQRSSRSHTIFFRMILESREKGE 217
Db      192 KKGVFVVGAVEQVVTSAEAYQVLGGWNRNRVASTSMNRESRSSHAVFTTIESMEK-- 249

QY      218 PSNCEGSKVSHNLNLDLAGSRAAQTGAAGVRLKEGCNINRSLFILGOVVKLSGQVVG 277
Db      250 -SNEIVNIRTSLLNLDLAGSERQKDTAEGMKLEAGNINRSLSCLGQVITALVD--VG 306

; CURRENT APPLICATION NUMBER: US/09/723,262
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Human
; US-09-723-262-6

Query Match          35.3%; Score 615; DB 4; Length 409;
Best Local Similarity 43.3%; Pred. No. 4.8e-57;
Matches 158; Conservative 48; Mismatches 113; Indels 46; Gaps 10;

QY      2 EGAIVAVCVVRPLNSREESL-GETAQVYWKTDNNVIYQVDGS-----KSFNFD 49
Db      23 EGDALKVVRIRPPAERSGSADGE-----QNCLSVLSSTSLRLHSNPEPKTFTFD 73

QY      50 RVFHGNETTKNVEEIAAIIIDSAIQYNGTIFAYGQTASGKTYTMMG-----SEDHL 102
Db      74 HVADVDVTTQESVFATVAKSIVESCMGNGYNGTIFAYGQTGSGKFTTMMGPSEDNFSHNL 133

QY      103 GVIPRAIHDF-----QKIKFPDPREFLLRVSYMEIYNETITDLCGTQKMKPLIREDV 157
Db      134 GVIPRSFEYLFSLIDREKEKAGAKSFLCKCSFIEIYNEQIYDLL--DSASAGLYLREHI 191

QY      158 NRVVYVADLTEEVVYVTSSEMALKWITKGEKSRHVGKTMNQRSSRSHTIFFRMILESREKGE 217
Db      192 KKGVFVVGAVEQVVTSAEAYQVLGGWNRNRVASTSMNRESRSSHAVFTTIESMEK-- 249

QY      218 PSNCEGSKVSHNLNLDLAGSRAAQTGAAGVRLKEGCNINRSLFILGOVVKLSGQVVG 277
Db      250 -SNEIVNIRTSLLNLDLAGSERQKDTAEGMKLEAGNINRSLSCLGQVITALVD--VG 306

; CURRENT APPLICATION NUMBER: US/09723219
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Human
; US-09-723-219-6

Query Match          35.3%; Score 615; DB 4; Length 409;
Best Local Similarity 43.3%; Pred. No. 4.8e-57;
Matches 158; Conservative 48; Mismatches 113; Indels 46; Gaps 10;

QY      2 EGAIVAVCVVRPLNSREESL-GETAQVYWKTDNNVIYQVDGS-----KSFNFD 49
Db      23 EGDALKVVRIRPPAERSGSADGE-----QNCLSVLSSTSLRLHSNPEPKTFTFD 73

QY      50 RVFHGNETTKNVEEIAAIIIDSAIQYNGTIFAYGQTASGKTYTMMG-----SEDHL 102
Db      74 HVADVDVTTQESVFATVAKSIVESCMGNGYNGTIFAYGQTGSGKFTTMMGPSEDNFSHNL 133

QY      103 GVIPRAIHDF-----QKIKFPDPREFLLRVSYMEIYNETITDLCGTQKMKPLIREDV 157
Db      134 GVIPRSFEYLFSLIDREKEKAGAKSFLCKCSFIEIYNEQIYDLL--DSASAGLYLREHI 191

QY      158 NRVVYVADLTEEVVYVTSSEMALKWITKGEKSRHVGKTMNQRSSRSHTIFFRMILESREKGE 217
Db      192 KKGVFVVGAVEQVVTSAEAYQVLGGWNRNRVASTSMNRESRSSHAVFTTIESMEK-- 249

QY      218 PSNCEGSKVSHNLNLDLAGSRAAQTGAAGVRLKEGCNINRSLFILGOVVKLSGQVVG 277
Db      250 -SNEIVNIRTSLLNLDLAGSERQKDTAEGMKLEAGNINRSLSCLGQVITALVD--VG 306

; CURRENT APPLICATION NUMBER: US/09723219
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Human
; US-09-723-219-6

Query Match          35.3%; Score 615; DB 4; Length 409;
Best Local Similarity 43.3%; Pred. No. 4.8e-57;
Matches 158; Conservative 48; Mismatches 113; Indels 46; Gaps 10;

QY      2 EGAIVAVCVVRPLNSREESL-GETAQVYWKTDNNVIYQVDGS-----KSFNFD 49
Db      23 EGDALKVVRIRPPAERSGSADGE-----QNCLSVLSSTSLRLHSNPEPKTFTFD 73

QY      50 RVFHGNETTKNVEEIAAIIIDSAIQYNGTIFAYGQTASGKTYTMMG-----SEDHL 102
Db      74 HVADVDVTTQESVFATVAKSIVESCMGNGYNGTIFAYGQTGSGKFTTMMGPSEDNFSHNL 133

QY      103 GVIPRAIHDF-----QKIKFPDPREFLLRVSYMEIYNETITDLCGTQKMKPLIREDV 157
Db      134 GVIPRSFEYLFSLIDREKEKAGAKSFLCKCSFIEIYNEQIYDLL--DSASAGLYLREHI 191

QY      158 NRVVYVADLTEEVVYVTSSEMALKWITKGEKSRHVGKTMNQRSSRSHTIFFRMILESREKGE 217
Db      192 KKGVFVVGAVEQVVTSAEAYQVLGGWNRNRVASTSMNRESRSSHAVFTTIESMEK-- 249

QY      218 PSNCEGSKVSHNLNLDLAGSRAAQTGAAGVRLKEGCNINRSLFILGOVVKLSGQVVG 277
Db      250 -SNEIVNIRTSLLNLDLAGSERQKDTAEGMKLEAGNINRSLSCLGQVITALVD--VG 306
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:39:47 ; Search time 16.9319 Seconds
(without alignments)
6280.361 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_340

Perfect score: 1744

Sequence: 1 AEEGAVAVCVRVPLNSREE.....QPASTAKYMKNTPYNEVST 339

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/ECTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09D_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09E_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	829.5	47.6	1382	16	US-10-437-963-176714	Sequence 176714, A
2	793.5	45.5	694	12	US-10-425-114-59725	Sequence 59725, A
3	668	38.3	1232	14	US-10-116-712-670	Sequence 670, App
4	668	38.3	1232	15	US-10-408-765A-2153	Sequence 2153, App
5	668	38.3	1235	15	US-10-334-143-8	Sequence 8, Appli
6	666	38.2	1006	16	US-10-437-963-147933	Sequence 147933, A
7	664	38.1	956	16	US-10-437-963-182113	Sequence 182113, A
8	663.5	38.0	955	12	US-10-425-114-62672	Sequence 62672, A
9	663	38.0	1237	15	US-10-334-143-33	Sequence 33, Appl
10	658	37.7	1232	14	US-10-116-712-664	Sequence 664, App
11	658	37.7	1232	14	US-10-116-712-669	Sequence 669, App
12	649.5	37.2	1029	16	US-10-311-642-2	Sequence 2, Appli
13	647.5	37.1	420	16	US-10-437-963-114373	Sequence 114373, A
14	630	36.1	672	16	US-10-408-765A-1664	Sequence 1664, Ap
15	622.5	35.7	965	16	US-10-437-963-169150	Sequence 169150, A

16	620.5	35.6	329	16	US-10-311-642-4	Sequence 4, Appli
17	619.5	35.5	966	12	US-10-425-114-62697	Sequence 62697, A
18	615.5	35.3	975	14	US-10-080-608A-19	Sequence 19, Appl
19	615.5	35.3	975	15	US-10-370-685-108	Sequence 108, App
20	615	35.3	409	12	US-10-332-089-6	Sequence 6, Appli
21	615	35.3	1388	12	US-10-332-089-2	Sequence 2, Appli
22	615	35.3	1388	14	US-10-146-473-82	Sequence 82, Appl
23	615	35.3	1388	15	US-10-173-999-32	Sequence 32, Appl
24	615	35.3	1388	16	US-10-188-832-164	Sequence 164, App
25	612.5	35.1	1031	14	US-10-080-608A-24	Sequence 24, Appl
26	612.5	35.1	1031	15	US-10-370-685-113	Sequence 113, App
27	609	34.9	1045	16	US-10-437-963-150217	Sequence 150217, A
28	605.5	34.7	1189	12	US-10-424-599-258016	Sequence 258016, A
29	604	34.6	963	14	US-10-080-608A-22	Sequence 22, Appl
30	604	34.6	963	15	US-10-370-685-111	Sequence 111, App
31	603.5	34.6	757	12	US-10-220-120-366	Sequence 366, App
32	603.5	34.6	757	16	US-10-363-829-316	Sequence 316, App
33	603	34.6	375	12	US-10-332-089-4	Sequence 4, Appli
34	603	34.6	963	14	US-10-080-608A-20	Sequence 20, Appl
35	603	34.6	963	15	US-10-370-685-109	Sequence 109, App
36	602.5	34.5	1011	12	US-10-336-472-24	Sequence 24, Appl
37	602.5	34.5	1011	12	US-10-236-417-58	Sequence 58, Appl
38	602.5	34.5	1032	14	US-10-080-608A-26	Sequence 26, Appl
39	602.5	34.5	1123	15	US-10-370-685-115	Sequence 115, App
40	602.5	34.5	1123	12	US-10-424-599-228690	Sequence 228690, A
41	600.5	34.4	1027	14	US-10-080-608A-27	Sequence 27, Appl
42	600.5	34.4	1027	15	US-10-370-685-116	Sequence 116, App
43	599	34.3	957	14	US-10-080-608A-16	Sequence 16, Appl
44	599	34.3	957	15	US-10-370-685-105	Sequence 105, App
45	598.5	34.3	517	15	US-10-369-493-12881	Sequence 12881, A

ALIGNMENTS

RESULT 1

US-10-437-963-176714
; Sequence 176714, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 176714
; LENGTH: 1382
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MBT4530_74437C.1.pap
US-10-437-963-176714

Query Match 47.6%; Score 829.5; DB 16; Length 1382;

Best Local Similarity 54.2%; Pred. No. 1.5e-75; Indels 17; Gaps 8;

Matches 186; Conservative 45; Mismatches 95;

QY 6 VAVCVRVPLNSREBSLGETAQVYKTDNNVT-YQVDGSKGFNFRVFGHNETTKNVEE 64

Db 4 IHVAVRPLTA--EDAGSSP---WRVSGNALALSTQPSIRPEFRIEGECRTADVCA 58

QY 65 TAAPIDSAIOGYNQTTIFAYGQTASGKTYTMWGSSEHIGVIPRAIHDFQTKKFPDREF 124

Db 59 RTKHIVDSAVRGFTGTFAYGQTNSTGKTYTWRGSGNEPGIIPLAVDLFRITIEHLDFREF 118


```
Query Match      38.3%; Score 668; DB 16; Length 1232;
Best Local Similarity 44.7%; Pred. No. 5.6e-59;
Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 5;

Qy 6 VAVCVVRPLNRRBSLG-ETAQVYWKTDNNVYQVDSKSFNFDRVFHGNHTTKNVEE 64
Db 10 VRVALRCRPLVPKEISEGCMCLSPVGPQVVGTD--KSTYDFVDPSTEQEVEVT 67

Qy 65 IAAPIIDSAIOGNGTIFAYGQTASGKTYTMWG-----SEDLGVIPRAIHDIFQKIK 117
Db 68 AVAPLIKGVFGYNATVLAAYGQTGSGKTYSMGAYTAQENEPYGVIPRVQLLFKEID 127

Qy 118 KPDPREFLLRVSYMEIYNETITDILCGTKMKPLIREDVNRNVYVADLTEEVVYVYTSMA 177
Db 128 KKSDFEFLTKVSYLEIYNEIIDLCPREKAQINREDPKREGIKVGLTEKTVLVALDT 187

Qy 178 LKWITKGSRHYGKTKMNRSSRSHITFIRMILESRKEGPEPNCSEGVKVSHLNVLVDLAG 237
Db 188 VSCLEQGNNSRTVASTAMNSQSSRSHAFTISLEQKKSD----KNSSFRSKLHLVLDLAG 243

Qy 238 SERAAQTGAAGVRLKEGNCINRSLFILGOVKKLSGQVGGFYNVYRDSKLTILQNSLGG 297
Db 244 SERQKKTAEGRDLKEGININRGLLCLGNVISALGDDKGGFVYRDSKLTILQNSLGG 303

Qy 298 NPKTRIICTITPV--SPDETITALQFASTAKYMKNTPTVYN 335
Db 304 NSHTLMIACVSPADSNLEETLTLAYADRARKIKKPIVN 343

RESULT 5
US-10-334-143-8
; Sequence 8, Application US/10334143
; Publication No. US20040009549A1
; GENERAL INFORMATION:
; APPLICANT: GRIGORIEV, IGOR VYACHESLAVOVICH
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: METHOD FOR DETECTING REMOTE HOMOLOGUES AND NOVEL
; TITLE OF INVENTION: KINASES IDENTIFIED WITH THE METHOD
; FILE REFERENCE: 038602/1543
; CURRENT APPLICATION NUMBER: US/10/334,143
; CURRENT FILING DATE: 2002-12-31
; PRIOR APPLICATION NUMBER: 60/343,169
; PRIOR FILING DATE: 2001-12-31
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 1235
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-334-143-8

Query Match      38.3%; Score 668; DB 15; Length 1235;
Best Local Similarity 44.7%; Pred. No. 5.6e-59;
Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 5;

Qy 6 VAVCVVRPLNRRBSLG-ETAQVYWKTDNNVYQVDSKSFNFDRVFHGNHTTKNVEE 64
Db 13 VRVALRCRPLVPKEISEGCMCLSPVGPQVVGTD--KSTYDFVDPSTEQEVEVT 70

Qy 65 IAAPIIDSAIOGNGTIFAYGQTASGKTYTMWG-----SEDLGVIPRAIHDIFQKIK 117
Db 71 AVAPLIKGVFGYNATVLAAYGQTGSGKTYSMGAYTAQENEPYGVIPRVQLLFKEID 130

Qy 118 KPDPREFLLRVSYMEIYNETITDILCGTKMKPLIREDVNRNVYVADLTEEVVYVYTSMA 177
Db 131 KKSDFEFLTKVSYLEIYNEIIDLCPREKAQINREDPKREGIKVGLTEKTVLVALDT 190

Qy 178 LKWITKGSRHYGKTKMNRSSRSHITFIRMILESRKEGPEPNCSEGVKVSHLNVLVDLAG 237
Db 191 VSCLEQGNNSRTVASTAMNSQSSRSHAFTISLEQKKSD----KNSSFRSKLHLVLDLAG 246

Qy 238 SERAAQTGAAGVRLKEGNCINRSLFILGOVKKLSGQVGGFYNVYRDSKLTILQNSLGG 297
Db 244 SERQKKTAEGRDLKEGININRGLLCLGNVISALGDDKGGFVYRDSKLTILQNSLGG 303
```

```
Db 247 SERQKKTAEGRDLKEGININRGLLCLGNVISALGDDKGGFVYRDSKLTILQNSLGG 306

Qy 298 NPKTRIICTITPV--SPDETITALQFASTAKYMKNTPTVYN 335
Db 307 NSHTLMIACVSPADSNLEETLTLAYADRARKIKKPIVN 346

RESULT 6
US-10-437-963-147933
; Sequence 147933, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 147933
; LENGTH: 1006
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1006)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MFT4530_48414C.1.pep
US-10-437-963-147933

Query Match      38.2%; Score 666; DB 16; Length 1006;
Best Local Similarity 44.9%; Pred. No. 6.6e-59;
Matches 154; Conservative 65; Mismatches 110; Indels 14; Gaps 9;

Qy 1 AEEGAVAVCVVRPLNRRBSLG-ETAQVYWKTDNNVYQVDSK-SFNFDRVFHGNHTTK 59
Db 64 AAKESVAVAVRPLSPREVRRGE--KIAWVADGETVARSEQSNLAYAYDRVFGPTTTTR 121

Qy 60 NYVEEIAAPIDSDAIOGNGTIFAYGQTASGKTYTMGSEDLGVIPRAIHDIFQKIKKF 119
Db 122 HIYDAVQVYVNGAMKNGTIFAYGVTSSGKTHTMGDIQISPGVPIPLAVKDFIWIQET 181

Qy 120 PDREFLLRVSYMEIYNETITDILCGTKMKPLIREDVNRNVYVADLTEEVVYVYTSMA 179
Db 182 PNREFLLRVSYLEIYNEVVDLL--NPAGONLRIBEDLQGTI-VEGIKEEAVLSPHALS 238

Qy 180 WITKGE---KSRHYGKTKMNRSSRSHITFIRMILESRKEGPEPNCSEGVKVSHLNVLDA 236
Db 239 LIAAGEVSELRHVSTNFNLLSSRSHITFTLTIESPRGQSEAE-AVTLSQLNLIDLA 297

Qy 237 GSERAAQTGAAGVRLKEGNCINRSLFILGOVKKLSGQVGGFYNVYRDSKLTILQNSLGG 296
Db 298 GSE-SSRYETAGVHQEGSYNKSLTLTGKVISKLT-D-EKATHIPFRDSKLTLLKSSLS 355

Qy 297 GNPKTRIICTITPVS--FDETLTALQFASTAKYMKNTPTVNEV 337
Db 356 GQGRVSLICTVTIPASSNSEETHNLKFAHRAKHIEIQATQNKI 398

RESULT 7
US-10-437-963-182113
; Sequence 182113, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
```

; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 182113
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(956)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_79331C.1.pep
; US-10-437-963-182113

Query Match 38.1%; Score 664; DB 16; Length 956;
Best Local Similarity 46.2%; Pred. No. 9.8e-59;
Matches 160; Conservative 52; Mismatches 112; Indels 22; Gaps 8;

Qy 2 BEGAVAVCVVRPLNSREESLGETAQVYWKDNN--VYQVD-----GSKSPNDRVPHG 54
Db 31 KEEKIFVTVRPLNSKKEALKD--QVAMECDNQTILYKPPQDRAAPTSTYFDKVFGP 88
Qy 55 NETTKNVEIEAIIIDSAIOGYNGTIFAYGQTASGKTYTMMGSEDLHGVIPRAIHDFQ 114
Db 89 ASQTEVVEEGAKDVAMSALTGINATIFAYGQTSSGKTFTMR-----GVTESAVNDIYR 142
Qy 115 KIKKPPDRFLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVYTS 174
Db 143 HIENTPERDFIIKISAMEIYNEIVKDLL--RPESTNRLDLPDPEKGTIVEKLEEEIAKDS 200
Qy 175 EMALKWITKGEKSRHYGETKMKQSRSSHTIFRMILESREKGEPSNCEGSKVYSHNLND 234
Db 201 QHLRLHSICEBQOVGETALNDTSRSHQIIRLTLESRLR-EVSGCVKSF-VANLNFVD 258
Qy 235 LAGSERAQTAAGVRLKEGNCINRSLFILGOVVKLSGQVGFINRDSKLTILQNS 294
Db 259 LAGSERAQTHAVGARLKEGCHINRSLTLTTVIRKLSGKRSCHIPYRDSKLTILQLS 318
Qy 295 LGGNPKTRIICTITPVSF---DEFTLQFALQFASAKYMKNTPYNEV 337
Db 319 LGGNARTAICTMSPAQTHVEQSRNTLFFATCAKEVTNNAKVMV 364

RESULT 8
US-10-425-114-62672
; Sequence 62672, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 62672
; LENGTH: 955

; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3689-241-A11_FLI.pep
; US-10-425-114-62672

Query Match 38.0%; Score 663.5; DB 12; Length 955;
Best Local Similarity 46.8%; Pred. No. 1.1e-58;
Matches 162; Conservative 46; Mismatches 117; Indels 21; Gaps 8;

Qy 1 AEEGAVAVCVVRPLNSREESLGETAQVYWK--TDNNVIYQVDS-----KSENFDRVEH 53
Db 33 AKEKIFVTVRPLNSKKE--LAAKDDVAMECDTQTLILYKPAQDRAAPMSYTFDKVFG 90
Qy 54 GNETTKNVEIEAIIIDSAIOGYNGTIFAYGQTASGKTYTMMGSEDLHGVIPRAIHDF 113
Db 91 PACQTDVVEEGAKDVAMSALTGINATIFAYGQTSSGKTFTMR-----GVTESAVSDIY 144
Qy 114 QKIKPPDRFLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVYTS 173
Db 145 RHIDSTPEREFVIKISAMEIYNEIVKDLL--RPDSAPFLRLDLPDPEKGTIVEKLEEEIAK 202
Qy 174 SEMALKWITKGEKSRHYGETKMKQSRSSHTIFRMILESREKGEPSNCEGSKVYSHNLV 233
Db 203 SOHLRLHSICEBQOVGETALNDTSRSHQIIRLTLESRLR-EVSGCVKSF-VANLNFV 260
Qy 234 LAGSERAQTAAGVRLKEGNCINRSLFILGOVVKLSGQVGFINRDSKLTILQNS 293
Db 261 LAGSERAQTHAVGARLKEGCHINRSLTLTTVIRKLSGKRSCHIPYRDSKLTILQLS 320
Qy 294 LGGNPKTRIICTITP--VSFDETLQFALQFASAKYMKNTPYNEV 337
Db 321 SLGGNARTAICTMSPALTHVEQSRNTLFFATCAKEVTNNAKVMV 366

RESULT 9

US-10-334-143-33
; Sequence 33, Application US/10334143
; Publication No. US20040009549A1
; GENERAL INFORMATION:
; APPLICANT: GRIGORIEV, IGOR VYACHESLAVOVICH
; APPLICANT: SUDASANAM, SUCHA
; TITLE OF INVENTION: METHOD FOR DETECTING REMOTE HOMOLOGUES AND NOVEL
; TITLE OF INVENTION: KINASES IDENTIFIED WITH THE METHOD
; FILE REFERENCE: 038602/1543
; CURRENT APPLICATION NUMBER: US/10/334,143
; PRIOR FILING DATE: 2002-12-31
; PRIOR FILING DATE: 2001-12-31
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 1237
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-334-143-33

Query Match 38.0%; Score 663; DB 15; Length 1237;
Best Local Similarity 44.1%; Pred. No. 1.9e-58;
Matches 150; Conservative 54; Mismatches 120; Indels 16; Gaps 6;

Qy 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSPNDRVPHGNETKNVYEE 64
Db 13 VRVALRCRPLVPKEISEGQCMCLSFVGETQVVGVD--KSFTYDFVDFDPTCEQEEVFNK 70
Qy 65 IAAPIIDSAIOGYNGTIFAYGQTASGKTYTMMG-----SEDLHGVIPRAIHDFQIK 117
Db 71 AVAPLKGIFKGNATVLAIGQTSKTYSMGAYTAQENEPVGIIPVLIQLFREID 130
Qy 118 KFPDRFLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVYTS 177
Db 131 KKSDFEFTLVSYLEIYNEEILDLLCPSRKAQINIREDPKEGIKIVGLTEKTVLVDLT 190

QY 178 LKWTGKSRHYGETKMNORSRSHITIFRMILESEKGEPSNCFDRVFGHNETTKNVEE 237
Db 191 VSCLEQGNNSRTVASTAMNSQSSRSHAITISIEQRKSD-KNCSFR--SKLHLVDLAG 246
QY 238 SERAAQTGAAGVRLKEGCNINSLFILGOVVKLSGQVGGFYNRYRDSKLTIRLQNSLGG 297
Db 247 SERQKTKAEGDRLKEGININRGLCLGNVISALGDDKKGFPAYRDSKLTIRLQNSLGG 306
QY 298 NPKTRIICITPV--SFDETLTALQFASAKYMNTPYV 335
Db 307 NSHTLMIACVSPADSNLEETLNTLRADYRARKIKNKPYN 346
RESULT 10
US-10-116-712-664
; Sequence 664, Application US/10116712
; Publication No. US20030194764A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Bangur, Chaitanya S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.568
; CURRENT APPLICATION NUMBER: US/10/116, 712
; CURRENT FILING DATE: 2002-04-07
; NUMBER OF SEQ ID NOS: 670
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 664
; LENGTH: 1232
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-116-712-664

Query Match 37.7%; Score 658; DB 14; Length 1232;
Best Local Similarity 44.7%; Pred. No. 6e-58;
Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 6;
QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDFRVFGHNETTKNVEE 64
Db 10 VRVALRCRLVPKEISEGCMCLSFVPGEPQVVVGTD--KSFTYDFVDFSTQEVEFNT 67
QY 65 IAAPIDSALQYNGTIFAYGQTASGKTYTMG-----SEDLGVIPRAIHDFOKIK 117
Db 68 AVAPLKGKGYGNATVLAAYGQTSKTYSMGAYTAQENETVGVIPRVQLLFEID 127
QY 118 KPDRFLLRVSMETINETITDLGCTOKMKPLIREDVNRNVYADLTEEYVYVSEMA 177
Db 128 KKSDFEFLKVSLEYLYNEBEILDLLCPSREKAQINIREDPKEGKIVGLTEKTVLVALDT 187
QY 178 LKWTGKSRHYGETKMNORSRSHITIFRMILESEKGEPSNCFDRVFGHNETTKNVEE 237
Db 188 VSCLEQGNNSRTVASTAMNSQSSRSHAITI---SLEQKSKDNSSFR-SKLHLVDLAG 243
QY 238 SERAAQTGAAGVRLKEGCNINSLFILGOVVKLSGQVGGFYNRYRDSKLTIRLQNSLGG 297
Db 244 SERQKTKAEGDRLKEGININRGLCLGNVISALGDDKKGFPAYRDSKLTIRLQNSLGG 303
QY 298 NPKTRIICITPV--SFDETLTALQFASAKYMNTPYV 335
Db 304 NSHTLMIACVSPADSNLEETLNTLRADYRARKIKNKPYN 343

RESULT 11
US-10-116-712-669
; Sequence 669, Application US/10116712
; Publication No. US20030194764A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Bangur, Chaitanya S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.568
; CURRENT APPLICATION NUMBER: US/10/116, 712

; CURRENT FILING DATE: 2002-04-07
; NUMBER OF SEQ ID NOS: 670
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 669
; LENGTH: 1232
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-116-712-669
Query Match 37.7%; Score 658; DB 14; Length 1232;
Best Local Similarity 44.7%; Pred. No. 6e-58;
Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 6;
QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDFRVFGHNETTKNVEE 64
Db 10 VRVALRCRLVPKEISEGCMCLSFVPGEPQVVVGTD--KSFTYDFVDFSTQEVEFNT 67
QY 65 IAAPIDSALQYNGTIFAYGQTASGKTYTMG-----SEDLGVIPRAIHDFOKIK 117
Db 68 AVAPLKGKGYGNATVLAAYGQTSKTYSMGAYTAQENETVGVIPRVQLLFEID 127
QY 118 KPDRFLLRVSMETINETITDLGCTOKMKPLIREDVNRNVYADLTEEYVYVSEMA 177
Db 128 KKSDFEFLKVSLEYLYNEBEILDLLCPSREKAQINIREDPKEGKIVGLTEKTVLVALDT 187
QY 178 LKWTGKSRHYGETKMNORSRSHITIFRMILESEKGEPSNCFDRVFGHNETTKNVEE 237
Db 188 VSCLEQGNNSRTVASTAMNSQSSRSHAITI---SLEQKSKDNSSFR-SKLHLVDLAG 243
QY 238 SERAAQTGAAGVRLKEGCNINSLFILGOVVKLSGQVGGFYNRYRDSKLTIRLQNSLGG 297
Db 244 SERQKTKAEGDRLKEGININRGLCLGNVISALGDDKKGFPAYRDSKLTIRLQNSLGG 303
QY 298 NPKTRIICITPV--SFDETLTALQFASAKYMNTPYV 335
Db 304 NSHTLMIACVSPADSNLEETLNTLRADYRARKIKNKPYN 343

RESULT 12
US-10-311-642-2
; Sequence 2, Application US/10311642
; Publication No. US2004008687A1
; GENERAL INFORMATION:
; APPLICANT: Cytokinetics, Inc.
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: NOVEL MOTOR PROTEINS AND METHODS FOR THEIR
; FILE REFERENCE: 020552-001910US
; CURRENT APPLICATION NUMBER: US/10/311,642
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: 09/597,602
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1029
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: HsKif17 amino acid sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HsKif17
US-10-311-642-2

Query Match 37.2%; Score 649.5; DB 16; Length 1029;
Best Local Similarity 42.8%; Pred. No. 3.4e-57;
Matches 149; Conservative 57; Mismatches 117; Indels 25; Gaps 8;
QY 5 AVAVCVVRPLNSREESLGETAQVYWKTD-----NNVIYQVDSKSFNDFRVFGHNE 56
Db 5 AVKVVVRCRPMNQREREL--RCQPVVTVDCARAQCCIQNPGAADPPKQFTFDGAIHVDH 62
QY 57 TTKNVYEEIAAPIIDSAIQGYNGTIFAYGQTASGKTYTMGSED--HLGVIPRAIHDF 113

Db 63 VTEQVNEIAYPLVEGTGNGTIFAYGTGSKSTWQCLPDPSPQGIIPFAFHFV 122
Qy 114 QKIKFPDPREFLLRVSMYNEITDILLCGTQKPKLIREDVNRNVYADLVEEYVT 173
Db 123 ESVQCAENTKPLVRASYLEIYNDVRLDGLADTKQK-LELKEHPEKGVYKGLSMHTVHS 181
Qy 174 SEMALKWITKGEKSRHGETKMQRSRSHITFIMILES-EGKSPNCEGSKVSHL 230
Db 182 VAQCEHMETQKNSRVGTYLNNKDSKSHSIFITSIEMSAVDERG-----KDLHLAGKL 236
Qy 231 NLVDLAGSERAQTCAGAVRLKEGNCINRSFILGOVVIKLSDGQGVGFINYRDSKLTRI 290
Db 237 NLVDLAGSERSQKGTATGERLKEATKINLSLALGNVISALVDGRC-KHVPYRDSKLTRL 295
Qy 291 LQNSLGNPKTRIITTPV--SPDETALQFASAKYMKNTPYVNE 336
Db 296 LQDSLGGNTKTLMAVCLSPADNNYDETLLSTLYANRAKNIRKPRINE 343
RESULT 13
US-10-437-963-114373
; Sequence 114373, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 114373
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_18070C.1.pcp
US-10-437-963-114373
Query Match 37.1%; Score 647.5; DB 16; Length 420;
Best Local Similarity 43.6%; Pred. No. 1.4e-57;
Matches 149; Conservative 57; Mismatches 115; Indels 21; Gaps 6;
Qy 6 VAVCVVRPLNREESLGHETAQYWKTDNNVIYQVDS-----KSFNDFRVHNET 57
Db 17 IVSVRLREPNAREAGDSD--WBCAGFTLTFRGAVPERAMFPASYSYDRVFSHECG 74
Qy 58 TKNVYEEIAPIDSAIQVNGTIFAYGTGSKTYTMMGSDHGLVPIRAIHDFQKIK 117
Db 75 TRQVDEGARQVAMSVLGINASIFAYGTSSKTYTMMGITEY-----SMSDIYDIE 128
Qy 118 KFPDREFLLRVSMYNEITDILLCGTQKPKLIREDVNRNVYADLVEEYVTSEMA 177
Db 129 KHPEREFILKFSAMEYNEAVRDL--SSDATPLRLDDPEKGTVVEKLTETLRDKGHL 186
Qy 178 LKWKITKGEKSRHGETKMQRSRSHITFIMILESREKGEPSNCEGSKVSHNLVDLAG 237
Db 187 LELAVCEAQROIGETAMNEASRSHOILRMVTESSAKQPLGKNSSTLIACVNFVDLAG 246
Qy 238 SERAAQTGAGVRLKEGNCINRSFILGOVVIKLSDGQGVGFINYRDSKLTRIQNSLGG 297
Db 247 SERASQTASAGMLKEGSHINRSLLTLGKVIQRLSKGR-NGHIPYRDSKLTRIQNSLGG 305
Qy 298 NPKTRIITCTTPV--SPDETALQFASAKYMKNTPYVNEV 337

Db 306 NARTAILCTMSPAHCHEIQSRNTLLFANCAKDVVTNAQVNVV 347
RESULT 14
US-10-408-765A-1664
; Sequence 1664, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Faby, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 650088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1664
; LENGTH: 672
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1664
Query Match 36.1%; Score 630; DB 16; Length 672;
Best Local Similarity 46.8%; Pred. No. 1.8e-55;
Matches 145; Conservative 49; Mismatches 102; Indels 14; Gaps 7;
Qy 37 IYQVDS----KSFNDFRVHNETKKNVYEEIAPIDSAIQVNGTIFAYGTGSKT 92
Db 18 VHKTDSSNEPKTFTFTVFGPESKQLDVYNTLPAIIDSVLGNGTIFAYGTGSKT 77
Qy 93 YTMGSE---DHLGVIPRAIHDFQKIKFP-DREFLLRVSMYNEITDILLCGTQK 148
Db 78 FTMEGVRAIPELGRGIIPNSFAHIFGHIAKAGDTFLVRVSVLEIYNEVRDL- GKDT 136
Qy 149 KPLIREDVNRNVYADLVEEYVTSEMAKWKITKGEKSRHGETKMQRSRSHITFIRM 208
Db 137 QLEVKERPVGVIKDLISAVVNNADDMDRIMTLGHKNRSVGCATNMNHSRSHAFIT 196
Qy 209 ILESREKGEPSNCEGSKVSHNLVDLAGSERAQTCAGAVRLKEGNCINRSFILGOV 268
Db 197 TTECSEKIDGNMH--VRMGKHLVDLAGSERQAKTGATGQRLKEATKINLSLTGNVI 254
Qy 269 KLSDGQGVGFINYRDSKLTRIQNSLGNPKTRIITTPV--SPDETALQFASAK 326
Db 255 SALVDGK-STHVPYRNSKLTRLQLQDSLGNSKTMCANIGPADYNYDETISTLYANRAK 313
Qy 327 YMKNTPYVNE 336
Db 314 NIKNKARINE 323
RESULT 15
US-10-437-963-169150
; Sequence 169150, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B

```

; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 169150
; LENGTH: 965
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_67597C.1.pap
US-10-437-963-169150

Query Match      35.7%; Score 622.5; DB 16; Length 965;
Best Local Similarity 42.5%; Pred. No. 1.9e-54;
Matches 147; Conservative 53; Mismatches 103; Indels 43; Gaps 7;

QY 1 AEEGAVAVCVVRPLNSRE-----ESLGETAQVYWKTDNNVIYQVDGSKSFNDRVFH 53
Db 17 AKEERIMVSVRLPLNGREAGDCDWCICISPTVMFRST---VPERAMFPTATYDRVFG 73

QY 54 GNETKNVVEETAAPIDSAIQYNGTIFAYGQTASGKTYTMMGSEDHLGVIPRAIHDF 113
Db 74 PDSSTRQVVEEGAKEVALSVVSGINSIFAYGQTSKGKTYTMTGITEY-----SVLDIY 127

QY 114 QKIKPPDPREFLLRVSYMEIYNETITDLLCGTQRMKPLIIREDVNRNVYVADLTEEVYT 173
Db 128 DYIEKHPEREFILRFSALIEIYNEAVRDLQ--SHDTTFLRLDDPEKGTTVKLTBTETLRD 185

QY 174 SEMALKWITKGEKSRHYGETKNQKRSRSHITFRMILESREKGEPSNCEGSKVSHLNV 233
Db 186 KDHLRNLAVCEAQRQIGETALNETSSRSHQILR-----LNFV 223

QY 234 DIAGSERAQTCGAAGVRLKEGNCINRSLFILGOVIKKLSDGOVGGFINYRDSKLTILON 293
Db 224 DIAGSERAQTCASAGVRLKEGSHINRSLTLTGKVVQKLSKGR--NGHIPRDSKLTILLOS 282

QY 294 SLGGNPKTRIICTITPV--SPDETTLAQFASTAKYMKNTPYNYEV 337
Db 283 SLGGNARTAIICTMSPARSHIEQSRNTLLFATCAKEVVTNAQNVV 328

```

Search completed: July 29, 2004, 10:06:25
Job time : 16.9319 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:28:21 ; Search time 5.0651 Seconds
(without alignments)
6437.961 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_340

Perfect score: 1744

Sequence: 1 ABEQAVAVCVVRPLNSREE.....QFASAKYMKNTPYNEVST 339

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1744	100.0	2563	1 S28261	centromere protein
2	1299.5	74.5	2954	2 T14156	kinesin-related pr
3	795.5	45.6	823	2 T52425	kinesin-like prote
4	760	43.6	888	2 D96619	protein T30E16.9 [
5	674	38.6	786	2 A53939	kinesin homolog KH
6	670	38.4	742	1 S58691	kinesin-related pr
7	666.5	38.2	1459	2 T30196	kinesin motor prot
8	665.5	38.2	1231	2 A54803	microtubule-associ
9	663	38.0	701	1 B44259	kinesin-related pr
10	659	37.8	747	1 A57107	kinesin-related pr
11	653.5	37.5	1226	2 T51617	kinesin-like prote
12	643.5	36.9	699	1 S38982	kinesin-related pr
13	633	36.3	672	2 S54351	kinesin oem-3 - Ca
14	633	36.3	932	2 T49235	kinesin-like prote
15	633	36.3	1225	2 A56514	chromokinesin - ch
16	631.5	36.2	909	2 H86350	hypothetical prote
17	621.5	35.6	1121	2 T06065	hypothetical prote
18	620.5	35.6	968	2 T45746	hypothetical prote
19	615.5	35.3	975	1 A31497	kinesin heavy chai
20	614	35.2	793	2 JC5831	kinesin-related pr
21	613	35.1	581	2 F84599	probable kinesin h
22	612.5	35.1	1031	1 A38713	kinesin heavy chai
23	612.5	35.1	1130	2 T21134	hypothetical prote
24	612	35.1	1076	2 B84687	probable kinesin-1
25	611	35.0	1254	2 T18277	kinesin heavy chai
26	609	34.9	1321	2 T13827	kinesin-73 - fruct
27	608.5	34.9	1263	2 T13465	hypothetical prote
28	606.5	34.8	294	2 S38983	kinesin-related pr
29	605	34.7	1066	1 A48669	kinesin-related pr

30	603.5	34.6	554	2 T50118	kinesin-related pr
31	603	34.6	963	1 A41919	kinesin heavy chai
32	602.5	34.5	1032	2 I38510	neuronal kinesin h
33	602.5	34.5	1388	2 T30335	KLP2 protein - Afr
34	598.5	34.3	967	1 A35075	kinesin heavy chai
35	598.5	34.3	1584	1 JN0114	kinesin-related pr
36	598.5	34.3	1584	2 T15822	kinesin-like prote
37	598	34.3	928	2 T10164	kinesin heavy chai
38	597	34.2	332	2 C48835	kinesin-like prote
39	595	34.1	968	2 T51933	kinesin motor prot
40	595	34.1	1027	2 S37711	kinesin heavy chai
41	591	33.9	843	2 S44868	kinesin heavy chai
42	588	33.7	784	1 A55236	kinesin-related pr
43	587.5	33.7	935	2 T51930	kinesin [imported]
44	586.5	33.6	1576	2 T29237	hypothetical prote
45	584	33.5	881	2 I84737	kinesin heavy chai

ALIGNMENTS

RESULT 1

S28261

centromere protein E - human

N/Alternate names: centromere 312K protein; kinesin-related protein CENP-E

C/Species: Homo sapiens (man)

C/Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001

C/Accession: S28261

R/Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.

Nature 359, 536-539, 1992

A/Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.

A/Reference number: S28261; MUID:93024922; PMID:1406971

A/Accession: S28261

A/Molecule type: mRNA

A/Residues: 1-2863 <YEN>

A/Cross-references: EMBL: Z15005; NID: g29864; PIDN: CAA78727.1; PID: g29865

C/Genetics:

A/Gene: GDB: CENPE

A/Cross-references: GDB: 361164; OMIM: 117143

A/Map position: 4q24-q25

C/Superfamily: centromere protein E; kinesin motor domain homology

C/Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop

F/7-335/Domain: kinesin motor domain homology <KMOT>

F/86-93/Region: nucleotide-binding motif A (P-loop)

F/486-2183/Domain: coiled coil #status predicted <COI>

F/92/Binding site: ATP (Lys) #status predicted

Query Match 100.0%; Score 1744; DB 1; Length 2663;
Best Local Similarity 100.0%; Pred. No. 5.6e-124;
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ABEGAVAVCVVRPLNSREESIGETAQVYKTDNNVYQVDCGSKSFNDRVPHGNETTKN	60
DB	2	ABEGAVAVCVVRPLNSREESIGETAQVYKTDNNVYQVDCGSKSFNDRVPHGNETTKN	61
QY	61	VVEETAAPIIDSAIQYNGTIFAYGQTASGKTYTWGSEDHLGVIPRAIHDIFQIKKFP	120
DB	62	VVEETAAPIIDSAIQYNGTIFAYGQTASGKTYTWGSEDHLGVIPRAIHDIFQIKKFP	121
QY	121	DREFLLRVSMEIYNETITDLCGTQMKPLIREDVNRNVYADLTTEVVYTSMALKW	180
DB	122	DREFLLRVSMEIYNETITDLCGTQMKPLIREDVNRNVYADLTTEVVYTSMALKW	181
QY	181	ITKGKSRHYGCTKMNQSRSHITFRMILSERKEGPNCEGSKVSHNLVDLAGSR	240
DB	182	ITKGKSRHYGCTKMNQSRSHITFRMILSERKEGPNCEGSKVSHNLVDLAGSR	241
QY	241	AAQTGAAGVRLKEGNCNINSLFILGVIKKISDGGVGGFINYRDNSKLTILQNSLGNPK	300
DB	242	AAQTGAAGVRLKEGNCNINSLFILGVIKKISDGGVGGFINYRDNSKLTILQNSLGNPK	301
QY	301	TRIICTITPVSFDELTALQFASTAKYMKNTPYNEVST	339

Db 302 TRIICITPVSFDETLTALOPASTAKYMNTPYVNEVST 340

RESULT 2

T14156

kinesin-related protein - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000

C:Accession: T14156

R:Wood, K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.

Cell 91, 357-366, 1997

A:Title: KENP-E is a plus end-directed kinetochore motor required for metaphase chromosome segregation

A:Reference number: Z17893; PMID:98028574; PMID:9363944

A:Accession: T14156

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2954 <WOO>

A:Cross-references: EMBL:AF027728; NID:g2586070; PID:g2586071; PIDN:AAC60300.1

C:Gene: KENP-E

C:Superfamily: centromere protein E; kinesin motor domain homology

Query Match 74.5%; Score 1299.5; DB 2; Length 2954;

Best Local Similarity 73.5%; Pred. No. 4.3e-90;

Matches 250; Conservative 42; Mismatches 43; Indels 5; Gaps 2;

Qy 1 ABEQAVAVCVVRPLNSREESLGETAQVYVTKTDNNVYQVDSKSFNDRVPHGNETTKN 60

Db 2 SEGDAVAVCVVRPLIQEQ--GDQANLQWAGNNTISQVDTGKSFNDRVFNESHESISQ 59

Qy 61 VYERIAAPIDSAIQGYNGTIPAYGOTASGKTYTMMGSEDLGVLTPRAIHDFQKIKFP 120

Db 60 IYOEIAVPIIRLSALQGYNGTIPAYGOTSSGKTYTMMGTPNSLGIIPQAIQEVFKLIQBP 119

Qy 121 DREFLLRYSMEIYNETITDLCGTQMKPLIREDVNRVYVADLTTEEYVYVTSMAWKW 180

Db 120 NREFLLRYSMEIYNETYKDLCDRRKKPLREDFNRVYVADLTTELWVPBHVQIW 179

Qy 181 ITGKSKRHYGETKNORSSRSHTIFRMILSRKGEPS--NCRGSKVVKVSHNLVLDLAG 237

Db 180 IKKGEKNRHYGETKNDRSSRSHTIFRMIVESDRNDNTSENCDGAVVSHNLVLDLAG 239

Qy 238 SERAAQTGAAGVRLKEGNCNINSLFILQGVIKKLSGQGVGGINYRDSKLTILQNSLGG 297

Db 240 SERASQTGAAGVRLKEGNCNINSLFILQGVIKKLSGQAGGFINYRDSKLTILQNSLGG 299

Qy 298 NPKTEIICITPVSFDETLTALOPASTAKYMNTPYVNEV 337

Db 300 NAKTVIICITPVSFDETLTALOPASTAKYMNTPYVNEV 339

RESULT 3

T52425

kinesin-like protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000

C:Accession: T52425

R:Kato, A.; Suzuki, M.; Kuwahara, A.; Ooe, H.; Higano-Inaba, K.; Komeda, Y.

Gene 239, 309-316, 1999

A:Title: Isolation and analysis of cDNA within a 300 kb Arabidopsis thaliana genomic region

A:Reference number: Z25171

A:Accession: T52425

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-823 <KAT>

A:Cross-references: EMBL:AB028468; PIDN:BAA88112.1

C:Gene: ZCF125

Query Match 45.6%; Score 795.5; DB 2; Length 823;

Best Local Similarity 51.4%; Pred. No. 1.7e-52;

Matches 178; Conservative 48; Mismatches 95; Indels 25; Gaps 7;

Qy 6 VAVCVVRPLNSREESLGETAQVYVTKTDNNVYQVDSKSFN-----FDRVPHGNE 56

Db 4 ICVAVRVRP-----PAPENGASLWKVEDN---RISLHKSLLDTPTTASHAFVDFDESS 54

Qy 57 TTKNVYBEIAAPIIDSAIQGYNGTIPAYGOTASGKTYTMMGSEDLGVLTPRAIHDFOKI 116

Db 55 TNASVYELLTKDIIHAIVEGFNGFAPAYGQISSKTFINTGSETDPIIRSVRDFERI 114

Qy 117 KKFDPDRFLRVSWEIYNETITDLCGTQMKPLIREDVNRVYVADLTTEEYVYVTSMAWK 176

Db 115 HMISDREFLLRVSWEIYNEEINLL--AVENORLQIHEHLRGVFAVLKEEIVSDAEQ 172

Qy 177 ALKWITGKSRHYGETKNORSSRSHTIFRMILSRKGEPSNCRGSKVVKVSHNLVLDLAG 236

Db 173 ILKLIDSGEVRNHFGETNMNVSRSHTIFRMVIESR--GKDNSSSDAIRVSVNLVLDLAG 230

Qy 237 GSERAAQTGAAGVRLKEGNCNINSLFILQGVIKKLSGQ--QVGGFINYRDSKLTILQNSL 295

Db 231 GSERIAKTGAGVRLQEGKYINKSLMILGNVINKLSGSTKLRAHPIYEDSKLTILQPAL 290

Qy 296 GGNPKTRIICITPVSV--FDETLTALOPASTAKYMNTPYVNEVST 339

Db 291 GGNAKTCIICITPAEEHHIEBSKGTLPQFASRAKRITNCAQVNEILT 336

RESULT 4

D96619

protein T30R16.9 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: D96619

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso-
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: AB6141; PMID:21016719; PMID:11130712

A:Accession: D96619

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-888 <STO>

A:Cross-references: GB:AB005173; NID:g8778739; PIDN:AAF79747.1; GSPDB:GN00141

C:Gene: T30R16.9

A:Map position: 1

Query Match 43.6%; Score 760; DB 2; Length 888;

Best Local Similarity 46.1%; Pred. No. 9.6e-50;

Matches 178; Conservative 48; Mismatches 96; Indels 62; Gaps 8;

Qy 6 VAVCVVRPLNSREESLGETAQVYVTKTDNNVYQVDSKSFN-----FDRVPHGNE 57

Db 4 ICVAVRVRP-----PAPENGASLWKVEDN---RISLHKSLLDTPTTASHAFVSGISITDLII 57

Qy 39 QVDSKSFN-----DRVPHGNETTKNVYEEIAAPIIDSAIQGYNGTIFAYGQTA 88

Db 58 EIVSLFLSFGVGVYFFLPADHVFDESSTNASVYELLTKDIIHAIVEGFNGTAFAYGQTS 117

Qy 89 SGKTYTMMGSEDLGVLTPRAIHDFQKIKFPDRFLRVSWEIYNETITDLCGTQMK 148

Db 118 SGKTFMTGSETDPIIRSVRDFERIHMISDRFLRVSWEIYNEEINLL--AVEN 175

Qy 149 KPLIREDVNRVYVADLTTEEYVYVTSMAWKWITGKSRHYGETKNORSSRSHTIFRM 208

Db 176 QRLQIHEHLRGVFAVLKEEIVSDAQILKLIDSGEVRNHFGETNMNVSRSHTIFRM 235

Qy 209 -----ILSRKGEPSNCRGSKVVKVSHNLVLDLAGSERAAQTGAAGVRLKEGNCN 256

[illegible]

Qy 87 TASGKTYPMGSEDLGVIPRAIHDIQIKKFPDREFLLRVSYMEIYNETITDLLCGTQ 146
Db 354 TSGKTHTMGSDAEPGVIPRAVEQIQMIKDEPREFLLRVSYMEIYNETIKDLIA-- 410
Qy 147 KKKPL-----IREDVNRNVYVADITEVW 171
Db 411 PLPLTGTSGSLQTTDRPASPIKGGSSHAAGSQSCCTLRIEDQKSRVITGLREEIV 470
Qy 172 YTSEMALKWITKGEKSRHVGTEKMNORSRSHITFRMILESRKGEPCNCEGSKVSHLN 231
Db 471 TDANTVLCIORQDERHVGAIDWNERSRSHCVFQLIESRSPASASKE--VRISQLN 528
Qy 232 IVDLAGSRAAQTGAAGVRLKEGCNINRSLFILGVQIKKLSGQVGG--FINYRDSKLT 289
Db 529 LIDLAGSRAA---SQAERRKEGAFINKSLTLGTIVIGKLTPEVNGDAHIPYRDSKLT 585
Qy 290 ILQNSLGNPKTRICTITPV--SFDETLTALQFASTAKYMKNT 331
Db 586 ILQTSLSGNARIAVICTLSPDTEHANETLSTLKFGKCKLVVTT 629
RESULT 8
A54803
N: microtubule-associated motor KIF4 - mouse
C: Species: Mus musculus (house mouse)
C: Date: 05-Apr-1995 #sequence_revision 05-Apr-1995 #text_change 02-Feb-2001
C: Accession: A54803; D44259
R: Sekine, Y.; Okada, Y.; Noda, Y.; Kondo, S.; Aizawa, H.; Takemura, R.; Hirokawa, N.
J. Cell Biol. 127, 187-201, 1994
A: Title: A novel microtubule-based motor protein (KIF4) for organelle transports, whose
A: Reference number: A54803; MUID: 95014709; PMID: 7923562
A: Accession: A54803
A: Molecule type: mRNA
A: Residues: 1-1231 <SEK>
A: Cross-references: GB:D12646; NID:9563772; PIDN: BAA02167.1; PID: d102657; PID: g563773
R: Aizawa, H.; Sekine, Y.; Takemura, R.; Zhang, Z.; Nangaku, M.; Hirokawa, N.
J. Cell Biol. 119, 1287-1296, 1992
A: Title: Kinesin family in murine central nervous system.
A: Reference number: A44259; MUID: 93077686; PMID: 1447303
A: Accession: D44259
A: Status: nucleic acid sequence not shown; not compared with conceptual translation
A: Molecule type: mRNA
A: Residues: 91-111, 'S', 113-240 <AIZ>
A: Experimental source: brain
A: Note: sequence extracted from NCBI backbone (NCBIP: 118904)
C: Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology
C: Keywords: ATP; nucleotide binding; P-loop
F: 10-343/Domain: kinesin motor domain homology <KNOT>
F: 88-95/Region: nucleotide-binding motif A (P-loop)

Query Match 38.2%; Score 665.5; DB 2; Length 1231;
Best Local Similarity 44.3%; Pred. No. 2.3e-42;
Matches 151; Conservative 55; Mismatches 118; Indels 17; Gaps 7;
Qy 6 VAVCVVRPLNSREESLG-ETAAQVYWKTDNNVYQVDSKSFNEDRVFPHGNETTKNVYEE 64
Db 10 VRVALRCRPLVSKKEICGQCTCLSFVGPGEQVW--VGNDKSFTYDFVDFDPSTQEVEPNT 67
Qy 65 IAAPLIDSAIQYNGTIFAYGQTASGKTYTMGWS-----EDH--LGVIPRAIHDFQKIK 117
Db 68 AVAPLIKGVFGYNATVLAIGTSGKTYSMGAYTAQEHDSATGIVPRVQLLFKEIN 127
Qy 118 KFPDREFLLRVSYMEIYNETITDLLCGT-QRMKPLIREDVNRNVYVADLTEEVYVTSEM 176
Db 128 KKSDFEFTLVKSYLFIYNEEILDLCSRRKATQINIREDPKEGKIIVGLETKTVLVASD 187
Qy 177 ALKWKITKGEKSRHVGTEKMNORSRSHITFRMILESRKGEPCNCEGSKVSHNLVDLA 236
Db 198 TVSCLEQGNNSRTVASTAMNSQSSRSRSHAFTISIEQRKK-----NDKNSFSRSLHLVDLA 243
Qy 237 GSERAAQTGAAGVRLKEGCNINRSLFILGVQIKKLSGQVGGFINYRDSKLTTRILQNSLG 296

Db 244 GSERQKTKABGRDLREGININRGLLCGNVISALGDDKGNFVYRDSKLTTRLLQDSL 303
Qy 297 GNPKTRIICTITPV--SFDETLTALQFASTAKYMKNTPEVN 335
Db 304 GNSHTLMIACVSPADSNLEETLNTLYRADRKIKNKPIIN 344
RESULT 9
B44259
N: kinesin-related protein KIF3A - mouse
C: Species: Mus musculus (house mouse)
C: Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C: Accession: B44259; S27872
R: Aizawa, H.; Sekine, Y.; Takemura, R.; Zhang, Z.; Nangaku, M.; Hirokawa, N.
J. Cell Biol. 119, 1287-1296, 1992
A: Title: Kinesin family in murine central nervous system.
A: Reference number: A44259; MUID: 93077686; PMID: 1447303
A: Accession: B44259
A: Molecule type: mRNA
A: Residues: 1-701 <AIZ>
A: Cross-references: EMBL:D12645; NID:9220469; PIDN: BAA02166.1; PID: g220470
A: Experimental source: brain
A: Note: sequence extracted from NCBI backbone (NCBIP: 118911)
C: Complex: heterodimer with KIF3B (PIR: A57107); the KIF3A/3B heterodimer associates with
C: Function:
A: Description: KIF3 complex is a motor protein that provides anterograde fast axonal tr
C: Superfamily: kinesin-related protein KIF3; kinesin motor domain homology
C: Keywords: ATP; coiled coil; heterodimer; heterotrimer; microtubule binding; nucleotid
F: 1-368/Domain: head globular #status predicted <HGI>
F: 15-354/Domain: kinesin motor domain homology <KNOT>
F: 100-107/Region: nucleotide-binding motif A (P-loop)
F: 369-599/Domain: helical rod #status predicted <ROD>
F: 600-701/Domain: tail globular #status predicted <TGL>
F: 106/Binding site: ATP (Lys) #status predicted

Query Match 38.0%; Score 663; DB 1; Length 701;
Best Local Similarity 46.1%; Pred. No. 1.7e-42;
Matches 158; Conservative 50; Mismatches 119; Indels 16; Gaps 8;
Qy 6 VAVCVVRPLNSREESLG--GETAAQVYWKTDNNVYQVDSG----KSFNEDRVFPHGNETTK 59
Db 15 VKVVRCPPLNERKSCMYRAQVSDMRGTITVHKTDSSNEPKTFTFDIVFGPESQKL 74
Qy 60 NVYBEIAAPLIDSAIQYNGTIFAYGQTASGKTYTMGSE---DHLGVIPRAIHDFQKI 116
Db 75 DVYNLTARPLIDSVLEGYNGTIFAYGQTGKTFTMEGVRAVPGRLRGVIPNSFAHIFGI 134
Qy 117 KKFP-DREFLLRVSYMEIYNETITDLLCGTQKMKPLIREDVNRNVYVADLTEEVYVTSE 175
Db 135 AKAEGRTRFLRVSYLFIYNEEVRDLL--GKDQTORLEVKERPDGVYIKDLSAYVNNAD 193
Qy 176 MALKWITKGEKSRHVGTEKMNORSRSHITFRMILESRKGEPCNCEGSKVSHNLVDL 235
Db 194 DMDRIMTLGHKVRSGVATNMNHSRSHAFTITIECKSGVDGNMH--VRMGKHLVDL 251
Qy 236 AGSERRAAQTGAAGVRLKEGCNINRSLFILGVQIKKLSGQVGGFINYRDSKLTTRILQNSL 295
Db 252 AGSERQAKTGATGQRLKEATKINLSLTLGNVISALVDGK--STHVPVNSKLTTRLLQDSL 310
Qy 296 GGNPKTRIICTITPV--SFDETLTALQFASTAKYMKNTPEYNE 336
Db 311 GGNSKTMWCANIGPADYNYDETITLTRYANRAKNKRNKARINE 353
RESULT 10
A57107
N: kinesin-related protein KIF3B - mouse
C: Species: Mus musculus (house mouse)
C: Date: 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change 19-Jan-2001
C: Accession: A57107
R: Yamazaki, H.; Nakata, T.; Okada, Y.; Hirokawa, N.
J. Cell Biol. 130, 1387-1399, 1995
A: Title: KIF3A/B: a heterodimeric kinesin superfamily protein that works as a microtubu


```

112 IFQKIKFFDP-BELLFVSVWEIYNETITDLIGTQXMKPLIITREDVNRNVVADLTTEV 170
Db      :      ||:||||:||||:||||:||||:||||:||||:||||:||||:
127 IFGHIAKEQENVRLFVSVSYLEIYNEVKOLL-GKQQQHLVKERPDVGVIYKDLSAFV 185
QY      :      ||:||||:||||:||||:||||:||||:||||:||||:||||:
171 VYTSMALKWTITKEGSRHVGETKQNRQSRSHITFRMILESRERKEGPFSCNCEGVSKVSHL 230
Db      :      ||:||||:||||:||||:||||:||||:||||:||||:||||:
186 VNNADDMDRIMTLGNKNRSVGATWNNESSRSHAIFITILERSDMG--LQXEQHVRVGKL 243
QY      :      ||:||||:||||:||||:||||:||||:||||:||||:||||:
231 NLVDLAGSRAAQTGAAGVRLKEGNCINRSLFILGQVIKKLSDGQGVGFINYRDSKLTRI 290
Db      :      ||:||||:||||:||||:||||:||||:||||:||||:||||:
244 HMVDLAGSROTKTGATGQRLKEATKINLSLSTLGNVISSLVGDK-STHPIYRNSKLTRL 302
QY      :      ||:||||:||||:||||:||||:||||:||||:||||:||||:
291 LQNSLGNPKTRIICITTFV--SPDETILALQFASTAKYMKNTPYVNE 336
Db      :      ||:||||:||||:||||:||||:||||:||||:||||:||||:
303 LQDSLGNAGNATVMCANIGPAEYNYDETISTLRYANRAKNIKNKAINE 350

RESULT 13
S54351
kinesin osm-3 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 02-Feb-2001
C/Accession: S54351
R/Tabish, M.; Siddiqui, Z.K.; Nishikawa, K.; Siddiqui, S.S.
J. Mol. Biol. 247, 377-389, 1995
A/Title: Exclusive expression of C. elegans osm-3 kinesin gene in chemosensory neurons
A/Reference number: S54351; MUID:95230679; PMID:7714894
A/Accession: S54351
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-672 <TAB>
A/Cross-references: GB:D38632; NID:g808864; PIDM:BAA07612.1; PID:d1008194; PID:gl490195
C/Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology
F/48-376/Domain: kinesin motor domain homology
C/Keywords: ATP; nucleotide binding; P-loop
F/130-137/Region: nucleotide-binding motif A (P-loop)

```

Query Match	36.3%	Score 533;	DB 2;	Length 672;
Best Local Similarity	47.0%;	Pred. No. 3e-40;		
Matches	142;	Conservative 49;	Mismatches 97;	Indels 14; Gaps 6;
Qy	41	DG-SKSFNEFRVPHGNETTKNVYEEIAAPIIDSALQYNGTIFAYGOTASGKTYTMMGSE	99	
Db	85	DGAARDFTDGAYPWRPGEQIYNDIVPFLVENVIEGYGVTFAYGQSGKTFSMQIE	144	
Qy	100	D---HLGVIPRAIHDIQKIKKFPDRBFLRVSYMEIYNETITDLCGTQRMKPLIRED	156	
Db	145	TIPAQRGVIPAFDHIPTATATTENVKFLVHCYLEIYNEVVDLLGADNKQK-LEIQEQ	203	
Qy	157	VNRNVVADLTVEVVYVYTEMALKWITGKBKSRHYGETKMNQORSRSHITPRMILLESREKQ	216	
Db	204	PDRGVYVAGLSMHVCHDVPACKERLMTGRFNRRHVGTALMNKDSRSHSIFTVYVEGITE-	262	
Qy	217	EPSNCEGSVKVSHLNLVDLAGSERAAQTGAAGVRLKEGCNINRSIFILGVIVKLSGQGV	276	
Db	263	-----TGSIRMGKLNLDLAGSERQSKTGATGDRLKEATKLNLSLALGNVISALVDGK-	316	
Qy	277	GGFINYRDSKLTIRLQNSLGNPKTRITCTTPVS--FDETLTALQFASATYKMNTPYV	334	
Db	317	SKHLPYRDSKLTIRLQDSLGNTKTIACVSPSSDNYDETLSTLRVYANRAKINKPPTI	376	
Qy	335	NE 336		
Db	377	NE 378		

RESULT 14
T49235
kinesin-like protein - Arabidopsis thaliana
N:Alternate names: protein F7K15.60
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49235

R;Obermaier, B.; Ottenwaelder, B.; Duchemin, D.; Zeitler, K.; Mewes, H.W.; Rudd, S.; submitted to the Protein Sequence Database, April 2000
A;Reference number: Z25019
A;Accession: T49235
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-932 <OBE>
A;Cross-references: EMBL:AL353871; GSPDB:GN00061; ATSP:F7KL5.60
A;Experimental source: cultivar Columbia; BAC clone F7KL5
C;Genetics:
A;Gene: ATSP:F7KL5.60
A;Map position: 3
A;Introns: 76/1; 106/1; 139/3; 177/3; 205/1; 227/3; 247/3; 289/2; 355/3; 401/3; 430/3;

Query Match	36.3%	Score 633;	DB 2;	Length 932;
Best Local Similarity	44.3%;	Pred. No. 4.7e-40;		
Matches	152;	Conservative 55;	Mismatches 116;	Indels 20;
Gaps	8			
Qy	3	EGAVAVCVRRPLNSREESLGETAQVYWK--TDNNVIYQ-----VDGSKSFNDRFVHGNE	56	
Db	26	EEKILVTVRPLNWRREHAKYDL--IAWCPDDEITIVFKNPENPKAPTKYGFDFKVFETC	83	
Qy	57	TTKNVYEEIAAPILDSIAQGYNGTIFAYGQTASGTATMMGSEDLGVIPRAIHDFOKI	116	
Db	84	ATQEVYEGGSDVALSALAGTNATIFAYGQTSSGKTFTMR-----GVTESVWKDIYEHI	137	
Qy	117	KKFPDRFLRVSMYIYNETITDILCGTQMKPLIREDVNRNVYVADLTVEEYVYTGEM	176	
Db	138	RKTOERSFVLKVSALYIYNETVVDLL--NRDTGPLRLDDPEKGTIVENLVEEYVESRQH	195	
Qy	177	ALKWITKGRSHRYGETKMKQRSSRSHITPEMILESREKGPSPNCEGVKVSHLMLVDLA	236	
Db	196	LQHLISICEQORQVGETALNDKSSRSHQIIRLTTHSSLR--ETAGCVQSF--MATLMLVDLA	253	
Qy	237	GSERAAQTGAAGVRLKEGCNNINRSLFILGQVIKKLSQGVGGFINRYSKULTRILQNSIG	296	
Db	254	GSERAFQTNADGLKLGSHNRSLLTITTVIRKLSGGRKDHVPYRYSKULTRILQNSIG	313	
Qy	297	GNPKTRILICTIPV--SFDFTLTALQFASTAKYMKNTPYVNEV	337	
Db	314	GNARTALICTISPAISHVEOTKTLTSFAMSAKEVNTCAKNVY	356	

RESULT 15
A56514
Chromokinesin - chicken
C:Species: Gallus gallus (chicken)
C:Date: 21-Jul-1995 #sequence_revision 28-Jul-1995 #text_change 02-Feb-2001
C:Accession: A56514; I50691
R;Wang, S.Z.; Adler, R.
J. Cell Biol. 128, 761-768, 1995
A:Title: Chromokinesin: a DNA-binding, kinesin-like nuclear protein.
A:Reference number: A56514; MUID:95181533; PMID:7876303
A:Accession: A56514
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1225 <WAN>
A:Cross-references: GB:U18309; NID:g603760; PIDN:AAC59666.1; PID:g603761
R;Wang, S.Z.; Adler, R.
Proc. Natl. Acad. Sci. U.S.A. 91, 1351-1355, 1994
A:Title: A developmentally regulated basic-leucine zipper-like gene and its expression
A:Reference number: A53451; MUID:94151328; PMID:8108415
A:Accession: I50691
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 728-1086,'RI' <WA2>
A:Cross-references: EMBL:U04821; NID:9440792; PIDN:AAA18960.1; PID:g440793
C:Genetics:

A;Gene: swi3-3
C;Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology
C;Keywords: ATP; DNA binding; mitosis; nucleotide binding; P-loop
F;11-344/Domain: kinesin motor domain homology <RMO>
F;88-95/Region: nucleotide-binding motif A (P-loop)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:26:36 ; Search time 3.25614 Seconds
(without alignments)
5421.082 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_340

Perfect score: 1744

Sequence: 1 AEEGAVAVCVRRVRLNLSREE.....QFASTAKYMKNTFYNEVST 339

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1744	100.0	2663	1	CENE_HUMAN	Q02224 homo sapien
2	674	38.6	786	1	FL10_CHLRE	P46869 chlamydomon
3	670	38.4	742	1	K121_STRPU	P46871 strongyloce
4	668	38.3	1232	1	KF4A_HUMAN	O95239 homo sapien
5	665.5	38.2	1231	1	KF4A_MOUSE	P33174 mus musculu
6	664	38.1	702	1	KF3A_HUMAN	Q9Y496 homo sapien
7	663	38.0	701	1	KF3A_MOUSE	P28741 mus musculu
8	660	37.8	747	1	KF3B_HUMAN	O15066 homo sapien
9	659	37.8	747	1	KF3B_MOUSE	Q61771 mus musculu
10	653.5	37.5	1226	1	KF4A_XENLA	Q91784 xenopus lae
11	649.5	37.2	1029	1	KF17_HUMAN	O9p2e2 homo sapien
12	643.5	36.9	699	1	K122_STRPU	P46872 strongyloce
13	641.5	36.8	1038	1	KF17_MOUSE	Q99pw8 mus musculu
14	633	36.3	672	1	OSM3_CAEEL	P46873 caenorhabdi
15	633	36.3	1225	1	KF4A_CHICK	Q90640 gallus gall
16	615.5	35.3	975	1	KINH_DROME	P17210 drosophila
17	614	35.2	793	1	KF3C_HUMAN	O14782 homo sapien
18	613	35.1	796	1	KF3C_RAT	O55165 rattus norv
19	612.5	35.1	1031	1	KINH_STRPU	P35978 strongyloce
20	611	35.0	796	1	KF3C_MOUSE	O35066 mus musculu
21	605	34.7	1066	1	KL61_DROME	P46863 drosophila
22	604	34.6	963	1	KINH_MOUSE	Q91768 mus musculu
23	603.5	34.6	554	1	KLP3_SCHPO	O9us60 schizosacch
24	603	34.6	963	1	KINH_HUMAN	P33176 homo sapien
25	602.5	34.5	1032	1	KINH_HUMAN	Q12840 homo sapien
26	602	34.5	1749	1	K13A_MOUSE	Q9eqw7 mus musculu
27	600.5	34.4	1027	1	KINH_MOUSE	P33175 mus musculu
28	599	34.3	957	1	KF5C_HUMAN	O60282 homo sapien
29	598.5	34.3	967	1	KINH_LOLPE	P21613 loligo peal
30	598.5	34.3	1584	1	U104_CAEEL	P23678 caenorhabdi
31	598	34.3	928	1	KINH_NEUCR	P48467 neurospora
32	598	34.3	1805	1	K13A_HUMAN	Q91rh9 homo sapien
33	595.5	34.1	1826	1	K13B_HUMAN	Q9hqt8 homo sapien

34	595	34.1	956	1	KF5C_MOUSE	P28738 mus musculu
35	591	33.9	815	1	KINH_CAEEL	P34540 caenorhabdi
36	588	33.7	784	1	KL68_DROME	P46867 drosophila
37	587.5	33.7	935	1	KINH_SYNRA	O43093 syncephalas
38	582.5	33.4	378	1	KLP2_BOMMO	P46874 bombyx mori
39	581	33.3	1690	1	KF1A_HUMAN	Q12756 homo sapien
40	579	33.2	1695	1	KF1A_MOUSE	P33173 mus musculu
41	576	33.0	1648	1	KF1A_HUMAN	Q15058 homo sapien
42	571	32.7	1816	1	KF1B_MOUSE	Q60575 mus musculu
43	562.5	32.3	1056	1	K125_ARATH	P82266 arabidopsis
44	559	32.1	1067	1	EG52_XENLA	Q91783 xenopus lae
45	558	32.0	1006	1	K125_TOBAC	O23826 nicotiana t

ALIGNMENTS

RESULT 1
CENE_HUMAN
ID CENE_HUMAN STANDARD; PRT; 2663 AA.
AC Q02224;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENPE
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis.";
RL Nature 359:536-539 (1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RL microtubule motor.";
RN [3]
RP EMBO J. 14:918-926 (1995).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RL interactions with the kinetochore proteins CENP-F and hBUBR1.";
RN [4]
RP FARNESYLATION.
RX MEDLINE=20459117; PubMed=10852915;
RA Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L.,
RA Bishop W.R., Kirschmeier P.;
RT "Farnesyl transferase inhibitors block the farnesylation of CENP-E
RT and CENP-F and alter the association of CENP-E with the
RT microtubules.";
RL J. Biol. Chem. 275:30451-30457 (2000).
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CC CONGRESSION. RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; Z15005; CAA78727.1; -.
CC DR MIM; L17143; -.
CC DR GO; GO:0005699; C:kinetochore; TAS.
CC DR GO; GO:0005634; C:nucleus; TAS.
CC DR GO; GO:0008350; F:kinetochore motor activity; TAS.
CC DR GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
CC DR GO; GO:0007079; P:mitotic chromosome movement; TAS.
CC DR GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
CC DR InterPro; IPR001752; Kinesin_motor.
CC DR Pfam; PF00225; kinesin; 1.
CC DR PRINTS; PR00380; kinesin; 1.
CC DR SMART; SM00129; KISC; 1.
CC DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
CC DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
CC KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
CC Cell cycle; Centromere; Lipoprotein; Prenylation.
CC FT DOMAIN 1 335 KINESIN-MOTOR.
CC FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
CC FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
CC FT NP_BIND 86 93 ATP (BY SIMILARITY).
CC FT LIPID 2660 2660 S-farnesyl cysteine.
CC SQ SEQUENCE 2663 AA; 312087 MW; CEFCL3880C8C8CB8 CRC64;

Query Match 100.0%; Score 1744; DB 1; Length 2663;
Best Local Similarity 100.0%; Pred. No. 9.3e-125;
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AEEGAVCVVRPLNRSREESLGETAQYVWKTNNVYQVDSKSFNDRVPHGNETKN 60
Db 2 AEEGAVCVVRPLNRSREESLGETAQYVWKTNNVYQVDSKSFNDRVPHGNETKN 61
Qy 61 VYEEIAPIIDSAIQYNGTTFAYCOTASGKTYTMGSEDLGVIPRAIHDFQIKKFP 120
Db 62 VYEEIAPIIDSAIQYNGTTFAYCOTASGKTYTMGSEDLGVIPRAIHDFQIKKFP 121
Qy 121 DRELLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVYVTSMAK 180
Db 122 DRELLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVYVTSMAK 181
Qy 181 ITKGEKSRHYGETKMNQSRSSHTIFRMILSRKGEPSNCEGSKVSHLNLVDLAGSER 240
Db 182 ITKGEKSRHYGETKMNQSRSSHTIFRMILSRKGEPSNCEGSKVSHLNLVDLAGSER 241
Qy 241 AAO7GAAGVRLKEGNCINRSFILGVTKLSDGVGGFINVRDSKLRILQNSLGGNPK 300
Db 242 AAO7GAAGVRLKEGNCINRSFILGVTKLSDGVGGFINVRDSKLRILQNSLGGNPK 301
Qy 301 TRIITCTIPVSPDETLTALQFASTAKYMKNTPYNEVST 339
Db 302 TRIITCTIPVSPDETLTALQFASTAKYMKNTPYNEVST 340

RESULT 2
FL10 CHLRE STANDARD; PRT; 786 AA.
AC P46869;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Kinesin-like protein FLA10 (KHP1 protein).
GN FLA10.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.

OX NCBI_TaxID=3055;
RN SEQUENCE FROM N.A.
RC STRAIN=137;
RX MEDLINE=94299638; PubMed=8027176;
RA Walther Z., Vashishtha M., Hall J.L.;
RT "the Chlamydomonas FLA10 gene encodes a novel kinesin-homologous
protein.";
RL J. Cell Biol. 126:175-188(1994).
CC -!- FUNCTION: Probably involved in flagellar assembly and maintenance.
CC May play a role in flagellar synthesis.
CC -!- TISSUE SPECIFICITY: Flagellar axoneme.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC II SUBFAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; L33697; AAA21738.1; -.
CC DR PIR; A53939; A53939.
CC DR HSP; P17119; 3KAR.
CC DR InterPro; IPR001752; kinesin_motor.
CC DR Pfam; PF00225; kinesin; 1.
CC DR PRINTS; PR00380; KINESINHEAVY.
CC DR SMART; SM00129; KISC; 1.
CC DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
CC DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
CC KW Motor protein; Microtubule; ATP-binding; Coiled coil.
CC FT DOMAIN 1 358 KINESIN-MOTOR (BY SIMILARITY).
CC FT DOMAIN 367 687 COILED COIL (POTENTIAL).
CC FT DOMAIN 688 786 GLOBULAR (POTENTIAL).
CC FT NP_BIND 97 104 ATP (POTENTIAL).
CC FT DOMAIN 388 391 POLY-GLY.
CC FT DOMAIN 705 714 POLY-GLY.
CC FT DOMAIN 756 759 POLY-ASP.
CC SQ SEQUENCE 786 AA; 86671 MW; F90969203EB79F1B CRC64;

Query Match 38.6%; Score 674; DB 1; Length 786;
Best Local Similarity 44.6%; Pred. No. 9e-44;
Matches 158; Conservative 59; Mismatches 113; Indels 24; Gaps 8;

Qy 5 AVAYCVVRPLNRSREESLGETAQYVWKTNNVY----QVDGS---KSFNDRVPHGNET 57
Db 10 SVKVVVRCPLNGKEKADGRSRIVDMVDVAGQVVRNPKADASEPPKAFDQVYDMNCQ 69
Qy 58 TKNVYEEIAPIIDSAIQYNGTTFAYCOTASGKTYTMGSEDLGVIPRAIHDFQ 114
Db 70 QRDVFDITARPLIDSCIEGYNGTTFAYCOTGTSHTMEGDEPELGLIPNFRYVFE 129
Qy 115 KI-VKFPDREPLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVYV 173
Db 130 IIAADSQTKFELVRSSVLEIYNEEVRLD--CKDHSKMKELKESPDGVYVVDLSQFVCKN 188
Qy 174 SEMALKWITTKGEKSRHYGETKMNQSRSSHTIFRMILSRK-----GEPSNCEGS 224
Db 189 YEEMNKVLLAGKNRQVGATLMNQDSRSRSHSIFTITIECIKLESAAAKQPKAKKDDSNH 248
Qy 225 VKVSHLNLVDLAGSERAAQTCAAGVRLKEGNCINRSFILGVTKLSDGVGGFINVRD 284
Db 249 VRVGKLNLDLAGSERQDKTGATGURLKEGKINLSLTALGNVISALVDGK-SGHIPYRD 307
Qy 285 SKLTRLQNSLGGNPKTRIICITIPV--SFDETLTALQFASTAKYMKNTPYNE 336
Db 308 SKLIFELLQDLSGGNTKTVMVANIGPADWNVDYDETMTSLRYANRAKNIQNKPKINE 361

RESULT 3
KI21_STRPU


```

DR InterPro: IPR001752; kinesin_motor.
DR Pfam: PF00225; kinesin: 1.
DR PRINTS: PR00380; KINESINHEAVY.
DR SMART: SM00129; KISC; 1.
DR PROSITE: PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE: PS0067; KINESIN MOTOR DOMAIN2; 1.
DR Motor protein; Microtubule; ATP-binding; DNA-binding;
KW Nuclear protein; Coiled coil.
FT DOMAIN 1 349 KINESIN-MOTOR.
FT DOMAIN 350 999 COILED COIL (BY SIMILARITY).
FT DOMAIN 1000 1232 GLOBULAR.
FT NP BIND 88 95 ATP (POTENTIAL).
FT CONFLICT 223 223 R -> G (IN REF. 2).
FT CONFLICT 231 231 S -> T (IN REF. 4).
FT CONFLICT 286 286 V -> A (IN REF. 2).
FT CONFLICT 422 422 L -> W (IN REF. 2).
FT CONFLICT 564 564 L -> H (IN REF. 4).
FT CONFLICT 564 564 L -> P (IN REF. 2).
FT CONFLICT 600 600 K -> E (IN REF. 3).
FT CONFLICT 668 668 R -> K (IN REF. 3 AND 4).
FT CONFLICT 928 928 Q -> P (IN REF. 1).
FT CONFLICT 958 958 Q -> R (IN REF. 3).
FT CONFLICT 960 960 L -> Q (IN REF. 1).
FT CONFLICT 996 997 LL -> S (IN REF. 4).
FT CONFLICT 1003 1014 QKHLPKDTLSP -> RTLPRIPFYQ (IN REF. 4).
FT CONFLICT 1022 1022 P -> Q (IN REF. 2).
FT CONFLICT 1077 1077 K -> N (IN REF. 2).
FT CONFLICT 1138 1138 G -> S (IN REF. 2).
FT CONFLICT 1232 AA; 139908 MW; FF74052A17A88F7 CRC64;
SQ SEQUENCE 1232 AA; 139908 MW; FF74052A17A88F7 CRC64;

Query Match 38.3%; Score 668; DB 1; Length 1232;
Best Local Similarity 44.3%; Pred. No. 4.8e-43;
Matches 151; Conservative 52; Mismatches 120; Indels 16; Gaps 5;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDRVPHGNETTKNVEE 64
DB 10 VRVALRCRPLVSEISGQCMLSFVPGEPOVTD--KSFTYDFVDFDSTEOEEVNT 67

QY 65 IAAPLIDSAIQYNGTIFAYGQTASGKTYTMWG-----SEDLGVIPRAIHDFOKIK 117
DB 68 AVAPLIXGVFGYNATVLAGYQGTSGKTYSMGGAYTAQEENEPVGVIPVLIQLFKXID 127

QY 118 KFPDFEFLRVSYMEIYNETITDLLCGTQKMKPLIREDVNRNVYVADLVEEVTSEMA 177
DB 128 KKSDFEFLKVSYLEIYNEELDLDCPSREKAQINIREDPKGIKIVGLTEKTVLVALDT 187

QY 178 LKWTGKESRHYGETKQNRSSRSHHTIFRMILSREKGEPSNCEGSKVYSHLNLVDLAG 237
DB 188 VSCLEQGNNSRTVASTAMNSQSSRSHAFITSLRQRKSD----KNSGFRSKHLVDLAG 243

QY 238 SERAAQTGAAGVRLKEGCNINRSLFILGQVVKLLSDGQVGFINYRDSKLTIRLQNSLGG 297
DB 244 SERQKTKAEGDRLKEGININEGLLGNVLSALGDDKKGQFVYRDSKLTIRLQDSLGG 303

QY 298 NPKTRIITCTPPV--SFDETLTALQFASAKYMNTPVYN 335
DB 304 NSHTLMIACVSPADSNLETLNTRYADRARKIKNKPTVN 343

RESULT 5
KF4A MOUSE STANDARD; PRT; 1231 AA.
AC KF4A MOUSE
DT 01-OCT-1993 (Rel. 27; Created)
DT 01-NOV-1995 (Rel. 32; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE Chromosome-associated kinesin KIF4A (Chromokinesin).
GN KIF4A OR KIF4 OR KNS4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]

```

```

RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
RC SPECIFICITY.
RC STRAIN=ICR; TISSUE=Brain;
RX MEDLINE=95014709; PubMed=7929562;
RA Sekine Y., Okada Y., Noda S., Aizawa H., Takemura R.,
RA Hirokawa N.;
RT "A novel microtubule-based motor protein (KIF4) for organelle
RT transports, whose expression is regulated developmentally.";
RL J. Cell Biol. 127:187-201(1994).
RN [2]
RP SEQUENCE OF 91-240 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93077686; PubMed=1447303;
RA Aizawa H., Sekine Y., Takemura R., Zhang Z., Nangaku M.,
RA Hirokawa N.;
RT "Kinesin family in murine central nervous system.";
RL J. Cell Biol. 119:1287-1296(1992).
CC -1- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR
CC -1- SPINDLE STABILIZATION.
CC -1- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic
CC chromosomes.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN PYRAMIDAL CELLS IN JUVENILE
CC HIPPOCAMPUS, GRANULAR CELLS IN JUVENILE CEREBELLAR CORTEX AND IN
CC ADULT SPLEEN.
CC -1- SIMILARITY: Belongs to the kinesin-like protein family.
CC Chromokinesin subfamily.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: D12646; BAA02167.1; --
CC PIR: A54803; A54803.
CC HSP; P17119; 3KAR.
CC MGD; MGI:108389; Kif4.
CC InterPro: IPR001752; kinesin_motor.
CC Pfam: PF00225; kinesin; 1.
CC PRINTS: PR00380; KINESINHEAVY.
CC SMART: SM00129; KISC; 1.
CC PROSITE: PS00411; KINESIN MOTOR DOMAIN1; 1.
CC PROSITE: PS0067; KINESIN MOTOR DOMAIN2; 1.
CC Motor protein; Microtubule; ATP-binding; DNA-binding;
KW Nuclear protein; Coiled coil.
FT DOMAIN 1 350 KINESIN-MOTOR.
FT DOMAIN 351 1000 COILED COIL (BY SIMILARITY).
FT DOMAIN 1001 1231 GLOBULAR.
FT NP BIND 88 95 ATP (POTENTIAL).
FT CONFLICT 112 112 I -> S (IN REF. 2).
FT CONFLICT 1231 AA; 139551 MW; F34F2C2D21158FE4 CRC64;
SQ SEQUENCE 1231 AA; 139551 MW; F34F2C2D21158FE4 CRC64;

Query Match 38.2%; Score 665.5; DB 1; Length 1231;
Best Local Similarity 44.3%; Pred. No. 7.4e-43;
Matches 151; Conservative 55; Mismatches 118; Indels 17; Gaps 7;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDRVPHGNETTKNVEE 64
DB 10 VRVALRCRPLVSEISGQCMLSFVPGEPOV--VGNDRSFTYDFVDFDSTEOEEVNT 67

QY 65 IAAPLIDSAIQYNGTIFAYGQTASGKTYTMWG-----EDH---LGVIPRAIHDFOKIK 117
DB 68 AVAPLIXGVFGYNATVLAGYQGTSGKTYSMGGAYTAQEENEPVGVIPVLIQLFKXIN 127

QY 118 KFPDFEFLRVSYMEIYNETITDLLCGTQKMKPLIREDVNRNVYVADLVEEVTSEMA 176
DB 128 KKSDFEFLKVSYLEIYNEELDLDCPSREKAQINIREDPKGIKIVGLTEKTVLNASD 187

QY 177 ALKWTGKESRHYGETKQNRSSRSHHTIFRMILSREKGEPSNCEGSKVYSHLNLVDLA 236
DB 188 TVSCLEQGNNSRTVASTAMNSQSSRSHAFITSLRQRK----NDKNSGFRSKHLVDLA 243

```



```

CC CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC CC II SUBFAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; AB002357; BAA20815.1; -
CC CC EMBL; AL121897; CAC16425.1; -
CC CC HSSP; P17119; 3KAR.
CC CC Genew; HGNC:6320; KIF3B.
CC CC MIM; 603754; -
CC CC GO; GO:0005873; C:plus-end kinesin complex; TAS.
CC CC GO; GO:0003777; F:microtubule motor activity; TAS.
CC CC GO; GO:0008574; F:plus-end-directed kinesin ATPase activity; TAS.
CC CC GO; GO:0008089; P:anterograde axon cargo transport; TAS.
CC CC GO; GO:0007368; P:determination of left/right asymmetry; TAS.
CC CC InterPro; IPR001752; kinesin_motor.
CC CC Pfam; PF00225; kinesin.
CC CC PRINTS; PR00380; KINESINHEAVY.
CC CC SMART; SM00129; KISC; 1.
CC CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
CC CC PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
CC CC Motor protein; Microtubule; ATP-binding; Coiled coil; Neurone.
CC CC KINESIN-MOTOR (BY SIMILARITY).
CC CC COILED COIL (BY SIMILARITY).
CC CC GLOBULAR.
CC CC NP_BIND 346 579
CC CC DOMAIN 580 747
CC CC FT DOMAIN 346 579
CC CC FT NP_BIND 96 103
CC CC FT DOMAIN 386 393
CC CC FT DOMAIN 394 406
CC CC FT DOMAIN 723 730
CC CC FT POLY-SER.
CC CC SEQUENCE 747 AA; 85125 MW; 97FA4573AFA87023 CRC64;

Query Match 37.8%; Score 660; DB 1; Length 747;
Best Local Similarity 44.2%; Pred. No. 9.9e-43;
Matches 153; Conservative 55; Mismatches 118; Indels 20; Gaps 7;

QY 5 AVAVCVVRPLNSRESLG-----ETAQVYWKTDNNVIYQVDSKSFNDFVRPHGN 55
DB 9 SVRVVVRCPMNGEKAASYDKVDVDVVKLGQSVKNPKGTSHM--PKTFTFDVAVDWN 66
QY 56 ETTKNVVEEIAPIIDSAIQYNGTIFAYGQTASGKTYTM---MGSEDLHGVIPRAIHD 112
DB 67 AKOFELYDETFRPLVDVSLQFNGTIFAYGQTGTGKTYTMEGIRGDPKRGVIPSFDHI 126
QY 113 FQIKKFPDRFLLRVSYMEIYNETITDLCGQTKMKPLIREDVNRNVVYADLTVEVY 172
DB 127 FTHISRSONQOYLVRASYLEIYQBEIRDLLSKDQ--TKRELKERPDTCGVYKDLSSFTVK 185
QY 173 TSEMALKWITKGEKSRHYGETKMKQSRSHITFRMILESRKGEKSPSCGSKVSVHNL 232
DB 186 SVKEIEHVMNVGNQNSRGATNMNEHSRSHAFVITIECEVSG--LDGENHVRVGLNL 243
QY 233 VDLAGSRAAQTAGAAGVRLKGCNINSLFTILGQVKKLSDQGVGFNYEDSKLTILO 292
DB 244 VDLAGSERQAKTAGGERLKEATKINLSLSALGNVISALVDGK--STHLPYRDSKLTILO 302
QY 293 NSLGGNPKTRIICTITPVP--DETLTALQASTAKYMKNTPPYNE 336
DB 303 DSLGNAKTVNVANVPASYNVEETLTTLTRVYANRAKINKKPRVNE 348

RESULT 9
KF3B MOUSE
ID KF3B MOUSE STANDARD; PRT; 747 AA.
AC Q61771;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```

```

DE DE Kinesin-like protein KIF3B (Microtubule plus end-directed kinesin
DE DE motor 3B).
DE GN KIF3B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Brain;
RX MEDLINE=96032268; PubMed=7559760;
RA Yamazaki H., Nakata T., Okada Y., Hirokawa N.;
RT "KIF3A/B: a heterodimeric kinesin superfamily protein that works as a
RT microtubule plus end-directed motor for membrane organelle
RT transport.";
RL J. Cell Biol. 130:1387-1399(1995).
CC -!- FUNCTION: Involved in tethering the chromosomes to the spindle
CC pole and in chromosome movement. Microtubule-based anterograde
CC translocator for membranous organelles. Plus end-directed
CC microtubule sliding activity in vitro (By similarity).
CC -!- SUBUNIT: Interacts with the SMC3 subunit of the cohesin
CC complex (By similarity). Heterodimer of KIF3A and KIF3B.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC II SUBFAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; D26077; BAA05070.1; -
CC CC PIR; A57107; A57107.
CC CC HSSP; P17119; 3KAR.
CC CC MGI; 107688; Kif3b.
CC CC InterPro; IPR001752; kinesin_motor.
CC CC Pfam; PF00225; kinesin; 1.
CC CC PRINTS; PR00380; KINESINHEAVY.
CC CC SMART; SM00129; KISC; 1.
CC CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
CC CC PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
CC CC Motor protein; Microtubule; ATP-binding; Coiled coil; Neurone.
CC CC KINESIN-MOTOR (BY SIMILARITY).
CC CC COILED COIL (BY SIMILARITY).
CC CC GLOBULAR.
CC CC NP_BIND 346 579
CC CC DOMAIN 580 747
CC CC FT DOMAIN 346 579
CC CC FT NP_BIND 96 103
CC CC FT DOMAIN 386 393
CC CC FT DOMAIN 394 405
CC CC FT POLY-GLU.
CC CC FT POLY-SER.
CC CC SEQUENCE 747 AA; 85288 MW; FA369AA130EC8B47 CRC64;

Query Match 37.8%; Score 659; DB 1; Length 747;
Best Local Similarity 44.2%; Pred. No. 1.2e-42;
Matches 153; Conservative 55; Mismatches 118; Indels 20; Gaps 7;

QY 5 AVAVCVVRPLNSRESLG-----ETAQVYWKTDNNVIYQVDSKSFNDFVRPHGN 55
DB 9 SVRVVVRCPMNGEKAASYDKVDVDVVKLGQSVKNPKGTSHM--PKTFTFDVAVDWN 66
QY 56 ETTKNVVEEIAPIIDSAIQYNGTIFAYGQTASGKTYTM---MGSEDLHGVIPRAIHD 112
DB 67 AKOFELYDETFRPLVDVSLQFNGTIFAYGQTGTGKTYTMEGIRGDPKRGVIPSFDHI 126
QY 113 FQIKKFPDRFLLRVSYMEIYNETITDLCGQTKMKPLIREDVNRNVVYADLTVEVY 172
DB 127 FTHISRSONQOYLVRASYLEIYQBEIRDLLSKDQ--TKRELKERPDTCGVYKDLSSFTVK 185
QY 173 TSEMALKWITKGEKSRHYGETKMKQSRSHITFRMILESRKGEKSPSCGSKVSVHNL 232
DB 186 SVKEIEHVMNVGNQNSRGATNMNEHSRSHAFVITIECEVSG--LDGENHVRVGLNL 243

```

```
QY 233 VDLAGSERAQAGVRKKECNINRSFILGQVKKLSDGQGVGFNRYRDSKLTIRLQ 292
Db 244 VDLAGSERAQAGVRKKECNINRSFILGQVKKLSDGQGVGFNRYRDSKLTIRLQ 302
QY 293 NSLGGNPKTRITICTIPVSF--DETTLALQFASAKYMKNTPYVNE 336
Db 303 DSLGGNAKIVWNVGPNVSYVEETLTTRYANRANKNKPRVNE 348

RESULT 10
KF4A XENLA
ID KF4A XENLA STANDARD; PRT; 1226 AA.
AC Q91784; Q9PS10.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chromosome-associated kinesin KLP1 (Chromokinesin).
GN KLP1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
GN [1]
RN SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
RP SPECIFICITY.
RC TISSUE=Oocyte;
RX MEDLINE=95236444; PubMed=7720067;
RA Vernos I., Raats J., Hirano T., Heasman J., Karsenti E., Wylie C.;
RT "Xklp1, a chromosomal kinesin-like protein essential for
spindle organization and chromosome positioning.";
RL Cell 81:117-127 (1995).
RN [2]
RN SEQUENCE OF 9-338 FROM N.A.
RX MEDLINE=93246065; PubMed=8482413;
RA Vernos I., Heasman J., Wylie C.;
RT "Multiple kinesin-like transcripts in Xenopus oocytes.";
RL Dev. Biol. 157:232-239 (1993).
CC -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR
SPINDLE STABILIZATION.
CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic
chromosomes.
CC -!- TISSUE SPECIFICITY: Expressed in oocytes, eggs, testes and
brain.
CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
Chromokinesin subfamily.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X82012; CAA57539.1; --
PIR; I51617; I51617.
HSP; P17119; 3KAR.
InterPro; IPR001752; kinesin_motor.
Pfam; PF00225; kinesin; 1.
PRINTS; PR00380; KINESINHEAVY.
SMART; SM00129; KISC; 1.
PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
KW Motor protein; Microtubule; ATP-binding; DNA-binding;
Nuclear protein; Coiled coil.
FT DOMAIN 1 350 KINESIN-MOTOR.
FT DOMAIN 351 1006 COILED COIL (BY SIMILARITY).
FT DOMAIN 1007 1226 GLOBULAR.
FT NP_BIND 87 94 ATP (POTENTIAL).
FT CONFLICT 163 163 I -> L (IN REF. 2).
SQ SEQUENCE 1226 AA; 138923 MW; 7F0275FCF3316697 CRC64;
```

```
Query Match 37.5%; Score 653.5; DB 1; Length 1226;
Best Local Similarity 44.4%; Pred. No. 6.1e-42;
Matches 154; Conservative 54; Mismatches 120; Indels 19; Gaps 9;

QY 2 BEG-AVAVCVRPVPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDRVHFHNETTK 59
Db 4 DEGIPVRVALRCFPLVPKNEGCKMCLTFVPGEQQVI--VGTEKSFTYDYVDFPSAQOE 61
QY 60 NYVEEIAAIPIDSAIQGYNGTIFAYQGTASGTYITMGSSEDH-----LGVIPRAIHDI 112
Db 62 EVVNSAVAPLIKGLFKGNATVLAIGQGTSGKTSYMGATYTHQNEPVTGVIPRIVIAL 121
QY 113 FQIKKFPDRELLRVSYMEIYNETITDLL-CGTQKMKPLIREDNRNRYVADLTTEVV 171
Db 122 FREIHQRPWEFNLKVSYLEITNEEILDLLIYAARDKNTISIREDEKGIKGLTERDV 181
QY 172 YTSEMALKWITKGEKSRHYGETKMNQSRSSHTIFRMILESRKGPSPNCEGSKVYSHLN 231
Db 182 KTAIDTSLCLEQGNSSRTVASTAMNQSQRSHAFITISIEQRKEGDKN---SFR-SKLH 237
QY 232 LVDLAGSERAQAGVRKKECNINRSFILGQVKKLSD-GQGVGFNRYRDSKLTIRI 290
Db 238 LVDLAGSERAQAGVRKKECNINRSFILGQVKKLSD-GQGVGFNRYRDSKLTIRL 297
QY 291 LQNSLGGNPKTRITICTIPV--SFDETTLALQFASAKYMKNTPYV 335
Db 298 LQNSLGGNPKTRITICTIPV--SFDETTLALQFASAKYMKNTPYV 344

RESULT 11
KF17_HUMAN
ID KF17_HUMAN STANDARD; PRT; 1029 AA.
AC Q9P2E2; Q95077; Q8N411;
DT 28-FEB-2003 (Rel. 41, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Kinesin-like protein KIF17 (KIF3-related motor protein).
GN KIF17 OR KIF3X OR KIAA1405.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
GN [1]
RN SEQUENCE FROM N.A.
RA Kimberley A., White S.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 78-1029 FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RA Chara O., Nagase T., Kikuno R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE OF 239-1029 FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVI.
The complete sequences of 150 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 7:65-73 (2000).
RN [4]
RN SEQUENCE OF 618-1029 FROM N.A. (ISOFORM 2).
RC TISSUE=Pancras;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
```


FT DOMAIN 1 340 KINESIN-MOTOR (BY SIMILARITY).
 FT DOMAIN 341 619 COILED COIL (BY SIMILARITY).
 FT DOMAIN 620 699 GLOBULAR (BY SIMILARITY).
 FT NP BIND 97 104 ATP (POTENTIAL).
 SQ SEQUENCE 699 AA; 78697 MW; 7B3866111CB08190 CRC64;
 Query Match 36.9%; Score 643.5; DB 1; Length 699;
 Best Local Similarity 46.0%; Pred. No. 1-6e-41;
 Matches 160; Conservative 49; Mismatches 114; Indels 25; Gaps 10;
 QY 6 VAVCVVRPLNREESLG-----ETAAVYWKTDNNVYQVDGSKSFNDFRVFHG 54
 Db 11 VRVVRCPPLNSKETGQFKSVVMDMRGTGV-----TNPNA-PSGEPPKSFTEFDFVFP 66
 QY 55 NETKNVYEIAPIIDSAIOGNGTIFAYGQTASGKTYTMMG--SEDHL-GVPIRAIHD 111
 Db 67 GAKQTDVYQNTARPIVDALIEGYNGTIFAYGQTGKTFTMEGVRGQPELRGIIPNSFAH 126
 QY 112 IFQIKKFPDR-EFLLRVSYMEIYNETITDLCCTQKMKPLIIRREDVNRNYYVADLTEEV 170
 Db 127 IFGHIAKEQENVFLRVSVLEIYNEEVKDLL-GKQOQHRLEVKERPVDGVVYVKDLISAFV 185
 QY 171 VYTSWALWKITGKSRHYGETKMNORSRSHITFRMILESEKGEPSNCEGSKVSHL 230
 Db 186 VNNADMDRIMTLGNKRSVCAATMNESSRSHAFTITILERSDMG--LDKEQHRVRGKL 243
 QY 231 NVDLAGSRAAQTGAAGVRLKEGCNINRSIFILGQVVKLSGQVGGFNYRDSKLTRI 290
 Db 244 HMVDLAGSERQTKTGATGQRLKEATKINLSLSTGNVSISSLVGK-STHPIYRNSKLTRL 302
 QY 291 LQSLGNPKTRIICITTPV--SFDETLTALQFASAKYMKNTPYVNE 336
 Db 303 LQSLGNNAKIVMCANIGPAEYNYDEITSLRYANRAKNIKAKINE 350
 RESULT 13
 KIF17 MOUSE STANDARD; PRT; 1038 AA.
 ID -KIF17 MOUSE
 AC Q99FW8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Kinesin-like protein KIF17 (MmKIF17).
 GN KIF17.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20307907; PubMed=10846156;
 RA Setou M., Nakagawa T., Seog D.-H., Hirokawa N.;
 RT "Kinesin superfamily motor protein KIF17 and mLin-10 in NMIIA
 RT receptor-containing vesicle transport.";
 RL Science 288:1796-1802(2000).
 CC -!- FUNCTION: Transports vesicles containing N-methyl-D-aspartate
 CC (NMDA) receptor 2B along microtubules.
 CC -!- SUBUNIT: Interacts with LIN-10 PDZ domain.
 CC -!- TISSUE SPECIFICITY: Neuronal-specific.
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; AB008867; BAB21099.1; -.
 CC DR HSSP; P17119; 3KAR.
 CC MGD; MGI:1098229; Kif17.

DR GO; GO:0005871; C:kinesin complex; IDA.
 DR GO; GO:0003777; F:microtubule motor activity; IDA.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0007017; P:microtubule-based process; IDA.
 DR GO; GO:0016192; P:vesicle-mediated transport; IDA.
 DR InterPro; IPR001752; Kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PRO0380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; Coiled coil; Transport;
 KW Protein transport. 265 KINESIN-MOTOR.
 FT DOMAIN 1 265 COILED COIL (POTENTIAL).
 FT DOMAIN 346 470 COILED COIL (POTENTIAL).
 FT DOMAIN 748 855 COILED COIL (POTENTIAL).
 FT NP BIND 91 98 ATP (POTENTIAL).
 SQ SEQUENCE 1038 AA; 116372 MW; 2BED852A3AFBD46 CRC64;
 Query Match 36.8%; Score 641.5; DB 1; Length 1038;
 Best Local Similarity 42.5%; Pred. No. 4e-41;
 Matches 147; Conservative 59; Mismatches 119; Indels 21; Gaps 8;
 QY 5 AVAVCVVRPLNREESLG-ETAAVYWKTDNNVYQVDG-----KSFNDFRVFHGNETT 58
 Db 5 SVKVVVRCRPMNKRERELSCQSVTVVDSARGQCFCIQNPGAADPPKQFTFDGAYYIEHFT 64
 QY 59 KNYVEIAPIIDSAIOGNGTIFAYGQTASGKTYTMMGSED--HLGVPIRAIHDIFOK 115
 Db 65 EQIYNEIAYPLVGVTEGYNGTIFAYGQTGSGSFTMQGLPDPQCGIIPRAFEHVFS 124
 QY 116 IKFPPDEFLRVSYMEIYNETITDLCGTQKMKPLIIRREDVNRNYYVADLTEEVYTS 175
 Db 125 VQCAENTKFLVRASYLEIYNEVDHLL-GADTKQRLKEHPEKGVVYVGLSMHTVHVA 183
 QY 176 MALKWITGKSRHYGETKMNORSRSHITFRMILE---SREKGEPSNCEGSKVSHLNL 232
 Db 184 QCERVMTGKNRAVGYTLMKDSRSSHSIFITNIEIYAVDERG-----KDHLAGKLN 238
 QY 233 VDLAGSRAAQTGAAGVRLKEGCNINRSIFILGQVVKLSGQVGGFNYRDSKLTRIQ 292
 Db 239 VDLAGSERQTKTGATGRLKEATKINLSLSTGNVSISSLVGRC-KHPIYRDSKLTRLQ 297
 QY 293 NSLGNPKTRIICITTPV--SFDETLTALQFASAKYMKNTPYVNE 336
 Db 298 DSLGNTKTLMVACLSPADNNYDEITSLRYANRAKNIKAKINEPRINE 343
 RESULT 14
 OSM3_CABEL STANDARD; PRT; 672 AA.
 ID -OSM3_CABEL
 AC P46873;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-1998 (Rel. 36, Last annotation update)
 DE Kinesin-like protein osm-3.
 GN OSM-3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=95230679; PubMed=7714894;
 RA Tabish M., Siddiqui Z.K., Nishikawa K., Siddiqui S.S.;
 RT "Exclusive expression of C. elegans osm-3 kinesin gene in
 RT chemosensory neurons open to the external environment.";
 RL J. Mol. Biol. 247:377-389(1995).
 RN [2]
 RP PRELIMINARY SEQUENCE OF 1-397 FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=93379214; PubMed=7690265;

RA Shahir M.A., Fukushige T., Yasuda H., Miwa J., Siddiqui S.S.;
 RT "C. elegans oem-3 gene mediating osmotic avoidance behaviour encodes
 RL a kinesin-like protein.";
 RL NeuroReport 4:891-894(1993).
 CC -!- TISSUE SPECIFICITY: AMPHID AND IL2 NEURONS.
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
 CC II SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D38632; BAA07612.1; -;
 CC EMBL; D14968; BAA20996.1; -;
 CC PIR; S54351; S54351.
 CC HSP; P17119; 3KAR.
 CC InterPro; IPR001752; kinesin_motor.
 CC PRINTS; PR00380; KINESINHEAVY.
 CC SMART; SM00129; KISC; 1.
 CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 CC PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
 CC Motor protein; Microtubule; ATP-binding; Coiled coil.
 CC DOMAIN 57 300 KINESIN-MOTOR.
 CC FT 461 497 COILED COIL (POTENTIAL).
 CC FT NP BIND 130 137 ATP (POTENTIAL).
 CC FT SEQUENCE 672 AA; 75456 MW; 482160F7042AFCD7 CRC64;
 CC
 CC Query Match 36.3%; Score 633; DB 1; Length 672;
 CC Best Local Similarity 47.0%; Pred. No. 9, 9e-41;
 CC Matches 142; Conservative 49; Mismatches 97; Indels 14; Gaps 6;
 CC
 CC 41 DG-SKSNFDRVFGNETKNVYVEIAPIIDSIAIQYNGTIFAYGQTASGKTYTMGSE 99
 CC 85 DGAADFTDFGAYPMIRPEQIYNDIVFPLVENIEGYNGTVFAYGQTSKTSMQGIE 144
 CC 100 D--HLGVPIRAIHDFQKIKKFPDRELLRVSYMEIYNEITDILCGTQKMKLPIRED 156
 CC 145 TIPQRGVIPAFPHITATATTENVFVHCSEIYNEVRLGADNKKQK-LEIKEQ 203
 CC 157 VNRNVYVADLVEEYVYVSEMALKWITKGEKSRHYGETKMNORSRSHITPRMILESREK 216
 CC 204 PDGRVYVAGLSMHVCHDVPACKELMTRGFNNRHVGATLMKDSRSRSHIFVYVEGITE- 262
 CC 217 EPSNCEGSVKVSHNLVDLAGSRAAQTGAAGVRLKEGCNINRSLFILGVYIKLSGQV 276
 CC 263 -----TGSIRMGKLNLDLAGSERQSKTGATGDRLEATKINLSALGNVISALVDGK- 316
 CC 277 GGFYNYRDSKLTRILQNSLGNPKTRICITPVY--FDETLTALQASTAKWNTKTPYV 334
 CC 317 SKHIPYRDSKLTRLQLSLGNTKTNIMACVSPSSDNYDETLSTLRVANKIKNPKPTI 376
 CC 335 NE 336
 CC 377 NE 378

RESULT 15

KF4A CHICK STANDARD; PRT; 1225 AA.
 AC Q90640; Q90608;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chromosome-associated kinesin KIF4A (Chromokinesin).
 GN KIF4A.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
 RP SPECIFICITY.
 RC STRAIN=White leghorn; TISSUE=Embryonic retina;
 RX MEDLINE=95181533; PubMed=7876303;
 RA Wang S.Z., Adler R.;
 RT "Chromokinesin: a DNA-binding, kinesin-like nuclear protein.";
 RL J. Cell Biol. 128:761-768(1995).
 RN [2]
 RP SEQUENCE OF 728-1088 FROM N.A.
 RC STRAIN=White leghorn; TISSUE=Embryonic retina;
 RX MEDLINE=94151328; PubMed=8108415;
 RA Wang S.Z., Adler R.;
 RT "A developmentally regulated basic-leucine zipper-like gene and its
 RT expression in embryonic retina and lens.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:1351-1355(1994).
 CC -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR
 CC SPINDLE STABILIZATION.
 CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic
 CC chromosomes.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN PROLIFERATING CELLS;
 CC NEUROEPITHELIUM OF EMBRYOS.
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
 CC Chromokinesin subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U18309; AAC59666.1; -;
 CC EMBL; U04821; AAA18960.1; -;
 CC PIR; A56514; A56514.
 CC HSP; P17119; 3KAR.
 CC InterPro; IPR001752; kinesin_motor.
 CC PRINTS; PR00380; KINESINHEAVY.
 CC SMART; SM00129; KISC; 1.
 CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 CC PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
 CC Motor protein; Microtubule; ATP-binding; DNA-binding;
 CC Nuclear protein; Coiled coil.
 CC FT DOMAIN 1 351 KINESIN-MOTOR.
 CC FT DOMAIN 352 1003 COILED COIL (BY SIMILARITY).
 CC FT DOMAIN 1004 1225 GLOBULAR.
 CC FT NP BIND 88 95 ATP (POTENTIAL).
 CC FT CONFLICT 1087 1088 KG -> RI (IN REF. 2).
 CC SQ SEQUENCE 1225 AA; 138923 MW; FA01ED83425F5875 CRC64;
 CC
 CC Query Match 36.3%; Score 633; DB 1; Length 1225;
 CC Best Local Similarity 42.8%; Pred. No. 2, 2e-40;
 CC Matches 148; Conservative 59; Mismatches 121; Indels 18; Gaps 8;
 CC
 CC 2 EGAVAVCVVRPLNRSREESLG-ETAQVYWKTDNNVIYQVDSKSNFDFVHGNETKN 60
 CC 6 EKGIPVRVVRCPRLVPKETSEGCQMCLSFVPGPQVI--VGSKATYDVDFPSVEQEE 63
 CC 61 VVEEIAAPIDSIAIQYNGTIFAYGQTASGKTYTMG-----SDH---LGVIPRAIHDF 113
 CC 64 VFTAVAPLIRGIFKGNATVLAAYGQTGSKTYSMGTVTASQEHDPMSGVIFRVIKLFF 123
 CC 114 QKIKFPDREFLRVSYMEIYNEITDILCGT-QKMKPLIREDVNRVNVVADLVEVY 172
 CC 124 KEKEQDWEFVLKVSYLEIYNEIDILDLCSRSRSQSISREDPKEGIKVGLTERNVA 183
 CC 173 TSEMAKWTITKGEKSRHYGETKMNORSRSHITPRMILESREKGEPSNCEGSVKVSHNL 232
 CC 184 SARDTVSCLEQNNCHTAVASTAWNSSQSSSHALFICIDQKKK----NDKNSFHSKJLHL 239

Fri Aug 6 10:49:09 2004

Qy	233	VDLAGSERAAQTGAAGVRLKEGCNINRSILFIILGOVKKL-SDQVGGFINYRDSKLTRIL	291
Db	240	VDLAGSERQKTKAEGDRLKEGININRGLLCLGNVISALGENKKGFPYRDSKLTRIL	299
Qy	292	QNSLGNPKTRIICTIPV--SFDETLTALQFASTAKYMKNTPYVN	335
Db	300	QDSLGNSHYTLMIACVSPADSNLEETLNTIRYADRARKIKKPIVN	345

Search completed: July 29, 2004, 09:36:14
Job time : 4.25614 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:27:31 ; Search time 14.3994 Seconds
(without alignments)
7428.155 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_340

Perfect score: 1744

Sequence: 1 AEEGAVAVCVVRPLNSREE.....QFASAKYMKNPYYNEVST 339

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPTREMBL 25.*

2: sp_archaea.*

3: sp_bacteria.*

4: sp_fungi.*

5: sp_human.*

6: sp_invertebrate.*

7: sp_mammal.*

8: sp_mhc.*

9: sp_organelle.*

10: sp_phage.*

11: sp_plant.*

12: sp_rodent.*

13: sp_virus.*

14: sp_vertebrate.*

15: sp_unclassified.*

16: sp_rvirus.*

17: sp_bacteriap.*

18: sp_bacteriap.*

19: sp_bacteriap.*

20: sp_bacteriap.*

21: sp_bacteriap.*

22: sp_bacteriap.*

23: sp_bacteriap.*

24: sp_bacteriap.*

25: sp_bacteriap.*

26: sp_bacteriap.*

27: sp_bacteriap.*

28: sp_bacteriap.*

29: sp_bacteriap.*

30: sp_bacteriap.*

31: sp_bacteriap.*

32: sp_bacteriap.*

33: sp_bacteriap.*

34: sp_bacteriap.*

35: sp_bacteriap.*

36: sp_bacteriap.*

37: sp_bacteriap.*

38: sp_bacteriap.*

39: sp_bacteriap.*

40: sp_bacteriap.*

41: sp_bacteriap.*

42: sp_bacteriap.*

43: sp_bacteriap.*

44: sp_bacteriap.*

45: sp_bacteriap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1538	88.2	549	11 Q7TPX4	Q7TPX4 mus musculus
2	1299.5	74.5	2954	13 Q42263	Q42263 xenopus lae
3	795.5	45.6	807	10 Q94HV9	Q94HV9 arabidopsis
4	795.5	45.6	823	10 Q9S7P3	Q9S7P3 arabidopsis
5	767.5	44.0	459	10 Q9S830	Q9S830 arabidopsis
6	760	43.6	888	10 Q9LQ62	Q9LQ62 arabidopsis
7	734	42.1	1885	5 Q869B8	Q869B8 dictyosteli
8	707	40.5	160	11 Q35059	Q35059 mus musculus
9	707	40.5	2013	5 Q9VK10	Q9VK10 drosophila
10	707	40.5	2244	5 Q9NCG0	Q9NCG0 drosophila
11	686	39.3	1055	10 Q8RW44	Q8RW44 arabidopsis
12	686	39.3	1055	10 Q8W5R5	Q8W5R5 arabidopsis
13	680	39.0	699	5 Q8MPT8	Q8MPT8 caenorhabdi
14	674	38.6	1058	10 Q9SJT0	Q9SJT0 arabidopsis
15	669.5	38.4	1033	10 Q9LHL9	Q9LHL9 arabidopsis
16	668	38.3	671	4 Q86XX7	Q86XX7 homo sapien

Q86tn3 homo sapien
Q9awm8 oryza sativ
P87198 ustilago ma
Q7tqg6 mus musculu
Q80yp3 mus musculu
Q8iwh8 homo sapien
Q80uk1 mus musculu
Q8cgj1 mus musculu
Q7tsz7 mus musculu
Q8s950 nicotiana t
Q8t11 xenopus lae
Q8bnh4 mus musculu
Q8ou27 mus musculu
Q8w5r6 arabidopsis
Q7yuc7 tetrahymena
Q9u0d5 tetrahymena
Q9c7b9 arabidopsis
Q8s905 arabidopsis
Q9vtrk drosophila
Q8mpt7 caenorhabdi
Q9gv93 caenorhabdi
Q19633 caenorhabdi
Q9gv92 caenorhabdi
Q802x4 brachydanio
Q9lxl3 arabidopsis
Q8lgu3 arabidopsis
Q815j2 arabidopsis
Q8lnz2 arabidopsis

ALIGNMENTS

RESULT 1

Q7TPX4 PRELIMINARY; PRT; 549 AA.

ID Q7TPX4
AC Q7TPX4
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Egg;
RX MEDLINE=22389257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heist P.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield V.S.,
Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
Jones S.J., Marra M.A.
"Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Egg;

RA Strausberg R.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC052843; AAHS2843.1; --
 KW Hypothetical protein.
 FT NON TER 549 549
 SQ SEQUENCE 549 AA; 62899 MW; B707C97DC9FB38D3 CRC64;

Query Match 88.2%; Score 1538; DB 11; Length 549;
 Best Local Similarity 88.2%; Pred. No. 3.6e-113;
 Matches 298; Conservative 11; Mismatches 29; Indels 0; Gaps 0;

QY 1 AEGAVAVCVVRPLNSREESLGETAQVYKTDNNVYQVDSKSFNDRVPHGNETTKN 60
 DB 1 AEGAVAVCVVRPLNSREESLGETAQVYKTDNNVYQVDSKSFNDRVPHGNETTKN 60
 QY 2 SEGDAVKCVVRPLIQREQ--GDQANLQWKAGNNTISQVDTGKSFNDRVPHGNETTSQ 59
 DB 2 SEGDAVKCVVRPLIQREQ--GDQANLQWKAGNNTISQVDTGKSFNDRVPHGNETTSQ 59
 QY 61 VYEETAAPIIDSAIQYNGTIFAYGQTASGKTYTMWGSDEHLGVIPRAIHDIFOKIKKFP 120
 DB 61 VYEETAAPIIDSAIQYNGTIFAYGQTASGKTYTMWGSDEHLGVIPRAIHDIFOKIKKFP 120
 QY 60 TYQETAVPIIRSAIQYNGTIFAYGQTSSGKTYTMWGTNSLGIIPQALQVFKIQEIP 119
 DB 60 TYQETAVPIIRSAIQYNGTIFAYGQTSSGKTYTMWGTNSLGIIPQALQVFKIQEIP 119
 QY 121 DREFLLRVSYMEIYNETITDLCGQTKMKPLIIRREDVNRNVYVADLTTEVVTSEMAKWK 180
 DB 121 DREFLLRVSYMEIYNETITDLCGQTKMKPLIIRREDVNRNVYVADLTTEVVTSEMAKWK 180
 QY 120 NREFLLRVSYMEIYNETVTKLLCDDRRKKPLEIREDFNENRVYVADLTTEELVMVPEVIOQ 179
 DB 120 NREFLLRVSYMEIYNETVTKLLCDDRRKKPLEIREDFNENRVYVADLTTEELVMVPEVIOQ 179
 QY 181 ITKGEKSRHYGETKMNQSRSSHTIFRMILESRKEGEPNCSGKSVKSHLNVDLAG 237
 DB 181 ITKGEKSRHYGETKMNQSRSSHTIFRMILESRKEGEPNCSGKSVKSHLNVDLAG 237
 QY 180 INKGEKSRHYGETKMNHSSRSHTIFRMIVESDRNDPTNSENCDGAVMWSHNLVDLAG 239
 DB 180 INKGEKSRHYGETKMNHSSRSHTIFRMIVESDRNDPTNSENCDGAVMWSHNLVDLAG 239
 QY 238 SERAOTGAAGVRLKEGCNINSLFVLGVKISDQVGGFINYRDSKLTIRLQNSLGG 297
 DB 238 SERAOTGAAGVRLKEGCNINSLFVLGVKISDQVGGFINYRDSKLTIRLQNSLGG 297
 QY 240 SERASQTGAAGVRLKEGCNINSLFVLGVKISDQVGGFINYRDSKLTIRLQNSLGG 299
 DB 240 SERASQTGAAGVRLKEGCNINSLFVLGVKISDQVGGFINYRDSKLTIRLQNSLGG 299
 QY 298 NPKTRIICTITPVSFDTLTALQFASAKYMKNTPYVNEV 337
 DB 300 NAKTVIICITPVSFDTLTALQFASAKYMKNTPYVNEV 339

RESULT 3
 Q94HV9 PRELIMINARY; PRT; 807 AA.
 ID Q94HV9
 AC Q94HV9;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Kinesin motor protein (kin2), putative.
 GN T4M14.11
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RC Town C.D., Haas B.J., Wu D., Maiti R., Hannick L.I., Chan A.P.,
 RA Tallon L.J., Rooney T., Utterback T.R., VanAken S.E., Feldblyum T.V.,
 RA White O., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome 1 BAC T4M14 genomic sequence."
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC027036; AAK62792.1; --
 DR GO; GO:0005871; C:kinesin complex; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003774; F:motor activity; IEA.
 DR GO; GO:0007017; P:microtubule-based process; IEA.
 DR InterPro; IPR001752; Kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
 SQ SEQUENCE 807 AA; 91260 MW; F07CDAD68EA30C28 CRC64;

Query Match 45.6%; Score 795.5; DB 10; Length 807;
 Best Local Similarity 51.4%; Pred. No. 3.7e-54;
 Matches 178; Conservative 48; Mismatches 95; Indels 25; Gaps 7;

QY 6 VAVCVVRPLNSREESLGETAQVYKTDNNVYQVDSKSFN-----PDRVPHGNE 56
 DB 4 ICVAVVRP-----PAPENGASLWKVEDN---RISLHKSLLDTPITASHAFDVFDESS 54
 QY 57 TTKNYVEEIAPIIDSAIQYNGTIFAYGQTASGKTYTMWGSDEHLGVIPRAIHDIFOKI 116
 DB 57 TTKNYVEEIAPIIDSAIQYNGTIFAYGQTASGKTYTMWGSDEHLGVIPRAIHDIFOKI 116

RA Strausberg R.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC052843; AAHS2843.1; --
 KW Hypothetical protein.
 FT NON TER 549 549
 SQ SEQUENCE 549 AA; 62899 MW; B707C97DC9FB38D3 CRC64;

Query Match 88.2%; Score 1538; DB 11; Length 549;
 Best Local Similarity 88.2%; Pred. No. 3.6e-113;
 Matches 298; Conservative 11; Mismatches 29; Indels 0; Gaps 0;

QY 1 AEGAVAVCVVRPLNSREESLGETAQVYKTDNNVYQVDSKSFNDRVPHGNETTKN 60
 DB 1 AEGAVAVCVVRPLNSREESLGETAQVYKTDNNVYQVDSKSFNDRVPHGNETTKN 60
 QY 2 SEGDAVKCVVRPLIQREQ--GDQANLQWKAGNNTISQVDTGKSFNDRVPHGNETTSQ 59
 DB 2 SEGDAVKCVVRPLIQREQ--GDQANLQWKAGNNTISQVDTGKSFNDRVPHGNETTSQ 59
 QY 61 VYEETAAPIIDSAIQYNGTIFAYGQTASGKTYTMWGSDEHLGVIPRAIHDIFOKIKKFP 120
 DB 61 VYEETAAPIIDSAIQYNGTIFAYGQTASGKTYTMWGSDEHLGVIPRAIHDIFOKIKKFP 120
 QY 62 VYEETAVPIIRSAIQYNGTIFAYGQTSSGKTYTMWGTNSLGIIPQALQVFKIQEIP 119
 DB 62 VYEETAVPIIRSAIQYNGTIFAYGQTSSGKTYTMWGTNSLGIIPQALQVFKIQEIP 119
 QY 121 DREFLLRVSYMEIYNETITDLCGQTKMKPLIIRREDVNRNVYVADLTTEVVTSEMAKWK 180
 DB 121 DREFLLRVSYMEIYNETITDLCGQTKMKPLIIRREDVNRNVYVADLTTEVVTSEMAKWK 180
 QY 122 EREFLLRVSYMEIYNETITDLCGQTKMKPLIIRREDVNRNVYVADLTTEVVTSEMAKWK 181
 DB 122 EREFLLRVSYMEIYNETITDLCGQTKMKPLIIRREDVNRNVYVADLTTEVVTSEMAKWK 181
 QY 181 ITKGEKSRHYGETKMNQSRSSHTIFRMILESRKEGEPNCSGKSVKSHLNVDLAG 237
 DB 181 ITKGEKSRHYGETKMNQSRSSHTIFRMILESRKEGEPNCSGKSVKSHLNVDLAG 237
 QY 182 LATGEKSRHYGETKMNQSRSSHTIFRMILESRKEGEPNCSGKSVKSHLNVDLAG 241
 DB 182 LATGEKSRHYGETKMNQSRSSHTIFRMILESRKEGEPNCSGKSVKSHLNVDLAG 241
 QY 241 AAGTGAAGVRLKEGCNINSLFVLGVKISDQVGGFINYRDSKLTIRLQNSLGGNPK 300
 DB 242 AAGTGAAGVRLKEGCNINSLFVLGVKISDQVGGFINYRDSKLTIRLQNSLGGNAX 301
 QY 301 TRIICITPVSFDTLTALQFASAKYMKNTPYVNEV 338
 DB 302 TRIICITPVSFDTLTALQFASAKYMKNTPYVNEV 339

RESULT 2
 Q42263 PRELIMINARY; PRT; 2954 AA.
 ID Q42263
 AC Q42263;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Kinesin-related protein.
 GN XENP-E. laevis (African clawed frog).
 OS Xenopus laevis; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Eukaryota; Metazoa; Chordata; Anura; Mesobatrachia; Pipidae;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 OC Xenopodinae; Xenopus.
 OC NCBI_TaxID=8355;
 RN SEQUENCE FROM N.A.
 RP MEDLINE=98028574; PubMed=9363944;
 RC Wood K.W., Sakowicz R., Goldstein L.S., Cleveland D.W.;
 RA "XENP-E is a plus end-directed kinetochore motor required for
 RT metaphase chromosome alignment."
 RL Cell 91:357-366 (1997).
 DR EMBL; AF027728; AAC60300.1; --
 DR PIR; T14156; 1BG2.
 DR HSSP; P33176; 1BG2.
 DR GO; GO:0005871; C:kinesin complex; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003774; F:motor activity; IEA.
 DR GO; GO:0007017; P:microtubule-based process; IEA.
 DR InterPro; IPR001752; Kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
 SQ SEQUENCE 2954 AA; 339964 MW; 439804ED0E592679 CRC64;

Query Match 74.5%; Score 1299.5; DB 13; Length 2954;

```

Db      55  TNASVELLT KDIIHAAVEGNGTAFAYGQTSSGKFTTMTGSETDPIIRSRVDVFERI 114
QY      117  KFPDPREFLLRVSYMEIYNETITDILCGTQKWKPLIIRDVNRNVYVADLTREVVYTSM 176
Db      115  HMSIDREFLLRVSYMEIYNEEINDLL--AVENQRLOIHEHLERGVPVAGLKEEIVSDAQ 172
QY      177  ALKWTITKESKSHYGETKMNQSSRSHTIFRMILSRKGEPSNCEGSKVSHLNLVDLA 236
Db      173  ILKLDGSEVNRHFGETNMNVHSSRSHTIFRMVIESR--GKDNSSSDAIRSVLNLVDLA 230
QY      237  GSERAQTCAGAVRLKGCNINRSFILGOVTKKLSDG--QVGGFINYRDSKLTIRLQNSL 295
Db      231  GSERAKTGAGVRLOEGKYINKSLMILGNVINKLSDSTKLAHTPYRDSKLTIRLQPAL 290
QY      296  GGNPKTRIICITPVS--FDELTALQFASHTAKYMKNTPYVNEVST 339
Db      291  GGNAKTCIICITPAEHHIESKGTLPFASRAKRTNCAQVNEILT 336.

RESULT 4
Q9S7P3 PRELIMINARY; PRT; 823 AA.
AC Q9S7P3
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-OCT-2003 (Tremblrel. 25, Last annotation update).
DE Kinesin-like protein.
GN ZCF125.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Ooe H., Kato A., Kameda Y.;
RT "Arabidopsis thaliana genomic sequence for a kinesin-like protein.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kato A., Suzuki M., Kuwahara A., Ooe H., Higano-Inaba K., Kameda Y.;
RT "Isolation and analysis of cDNA within a 300 kb Arabidopsis thaliana
RL genomic region located around the 100 map unit of chromosome 1.";
RL Gene 239:309-316(1999).
DR EMBL; AB028470; BAA88114.1; -
DR EMBL; AB028468; BAA88112.1; -
DR PIR; T52425; T52425.
DR HSSP; P33176; 1BG2.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IPR001752; Kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
SQ SEQUENCE 823 AA; 93148 MW; 6AFB1C622E4632C9 CRC64;

Query Match 45.6%; Score 795.5; DB 10; Length 823;
Best Local Similarity 51.4%; Pred. No. 3.8e-54;
Matches 178; Conservative 48; Mismatches 95; Indels 25; Gaps 7;

QY      6  VAVCVVRPLNRSREISGETAQYVWTKDNNVYQVDSKSFN-----FDRVHGNE 56
Db      4  ICVAVRVP-----PAPENGASLWKVEDN---RISLHKSLDTPITASHAFDHFVDESS 54
QY      57  TTKNVVEEIAAPLIDSAIQYNGTIFAYGQTSSGKFTTMTGSETDPIIRSRVDVFERI 116

```

```

Db      55  TNASVELLT KDIIHAAVEGNGTAFAYGQTSSGKFTTMTGSETDPIIRSRVDVFERI 114
QY      117  KFPDPREFLLRVSYMEIYNETITDILCGTQKWKPLIIRDVNRNVYVADLTREVVYTSM 176
Db      115  HMSIDREFLLRVSYMEIYNEEINDLL--AVENQRLOIHEHLERGVPVAGLKEEIVSDAQ 172
QY      177  ALKWTITKESKSHYGETKMNQSSRSHTIFRMILSRKGEPSNCEGSKVSHLNLVDLA 236
Db      173  ILKLDGSEVNRHFGETNMNVHSSRSHTIFRMVIESR--GKDNSSSDAIRSVLNLVDLA 230
QY      237  GSERAQTCAGAVRLKGCNINRSFILGOVTKKLSDG--QVGGFINYRDSKLTIRLQNSL 295
Db      231  GSERAKTGAGVRLOEGKYINKSLMILGNVINKLSDSTKLAHTPYRDSKLTIRLQPAL 290
QY      296  GGNPKTRIICITPVS--FDELTALQFASHTAKYMKNTPYVNEVST 339
Db      291  GGNAKTCIICITPAEHHIESKGTLPFASRAKRTNCAQVNEILT 336.

RESULT 5
Q9S830 PRELIMINARY; PRT; 459 AA.
AC Q9S830
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Putative kinesin-like centromere protein.
GN F14P13.22.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Uterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC F14P13 genomic sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC009400; AAF02823.1; -.
DR HSSP; P33176; 1BG2.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IPR001752; Kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
SQ SEQUENCE 459 AA; 51872 MW; 7C8487E9B7038E6A CRC64;

Query Match 44.0%; Score 767.5; DB 10; Length 459;
Best Local Similarity 55.5%; Pred. No. 2.7e-52;
Matches 166; Conservative 45; Mismatches 77; Indels 11; Gaps 5;

QY      49  DRVFHGNETTKNVYEIAAPIIDSAIQYNGTTFAYGQTSSGKFTTMTGSETDPIIRSRVDVFERI 108
Db      45  DRIFREDCKTVQVYEARTKEIVSAARGVNGTVFAIGQTNSTGKTHTRGSPIEPGVPLA 104
QY      109  IHDIPOKIKKPPDREFLLRVSYMEIYNETITDILCGTQKWKPLIIRDVNRNVYVADLTRE 168
Db      105  VHLDFDTIYQDASREFLLRMSYLEIYNEDINDLL--APEHRKLIQHLENLEKGFVAGLRE 162
QY      169  EVVYTTSEMALKWITKESKSHYGETKMNQSSRSHTIFRMILSRKGEPSNCEGSKVSHLNLVDLA 224
Db      163  EIVASPPQVLEWMEFGESEHRIHGETNMNLYSSRSHTIFRMILSRKGEPSNCEGSKVSHLNLVDLA 221
QY      225  VKVSHLNLVDLAGSRAAQTGAAGVRLKGCNINRSFILGOVTKKLSDG--QVGGFINY 292

```


Matches	166;	Conservative	51;	Mismatches	103;	Indels	16;	Gaps	8;
QY	5	AVAVCVVRPLNSREESLGETAQVYKNT-DNNVIYQVDG-SKSFNFDYRPHGNETTKNYV	62						
Db	8	SIQVCIKVRPCPLISL-----WQVKERRSIHLADSHAPYVDFVDEGASMQEVE	60						
QY	63	EELAAPILDISAIGYNGTIPAYGQTASGKTYTWMGSEDLGVTPRAIHDFIKIKFPDR	122						
Db	61	DRMAKHIVHACMQGNGTIFAYGQTSKGKTYTWMGDEQNPGVMVLAKEIFOQISSETER	120						
QY	123	EFLIRSVMEIYNETITDLLCGTKQMKPLIREDVNRNVVADITEEVVYTSWM-ALKWI	181						
Db	121	DFLLRVGIYIYNEKIYDILL--NKNQDLKHESGNGIVNVN--CBECIITSEVDLLRL	176						
QY	182	TKGEKSHRYGETKMNORSSRSHTIFRMILESRKGPESNCEGSKYVSHLNLVDLAGSERA	241						
Db	177	CLGNKERTVGETNNERSRSHAFKIIIESR-KSDHSD-DDAVIQSVNLVDLAGSERA	234						
QY	242	AQTGAAGVRLKEGNCINRSLFLQGVTKKLSDGQVGGFINYRDSKLTIRILQNSLGNPKT	301						
Db	235	DQTGARGARLKEGGHINKSLFLSNVTKLSLNADNRFNYRDSKLTIRILQSLGGNAFT	294						
QY	302	RIICITPVSFDELTALQFASTAKYMKNTPYNNEV	337						
Db	295	SIICITKPSIMEESQSTLSFATRAKKIRIKPQVNM	330						

RESULT 10

Q9NCGO PRELIMINARY; PRT: 2244 AA.

AC Q9NCGO;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Kinesin-like kinetochore motor protein CENP-meta.

GN CMET OR CG6392.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20351410; PubMed=10893249;

RA Yucel J.K., Marszalek J.D., McIntosh J.R., Goldstein L.S.B., Cleveland D.W., Philip A.V.;

RT "CENP-meta, an Essential Kinetochore Kinesin Required for the Maintenance of Metaphase Chromosome Alignment in Drosophila.";

RL J. Cell Biol. 150:1-12(2000).

DR EMBL; AF220353; AAF32355.1; -.

DR HSSP; P31176; 1BG2.

DR FlyBase; FBgn0040232; cmet.

DR GO; GO:0005699; C:kinetochore; IDA.

DR GO; GO:0007080; P:mitotic metaphase plate congression; IMP.

DR InterPro; IPR001752; kinesin_motor.

DR Pfam; PF00225; kinesin; 1.

DR PRINTS; PR00380; KINESINHEAVY.

DR SMART; SM00129; KISC; 1.

DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.

DR PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.

SO SEQUENCE 2244 AA; 257999 MW; FA6AA3B2A541ADE80 CRC64;

	Query Match	40.5%; Score 707; DB 5; Length 2244;	
	Best Local Similarity	49.4%; Pred. No. 1.7e-46;	
	Matches 166; Conservative	51; Mismatches 103; Indels 16; Gaps	8
Qy	5	AVACVVRPLMSREESIGETAQYWK-TDNNVIQVDG-SKSPNDEVRPHGNETTKNYV	62
Dd	8	SIVQIKVRPCBPGLTSL-----WQVKEGRSIIHLADSHAEPYFYVDDEGASNQEVF	60
Qy	63	EIEAIPIDSAIQGYNGIFAYGQTASCKTYTMGSEDHILGVIPRAIHDIFQIKKKPPDR	122
Dd	61	DMRARIHVACMQGFNGITFIAYGQTSKCTTYTMGDQNPGVMYLAAKEQQOISSETER	120

QY	123	PELLRVSYMEIYNETITDLCCTGCKMPLIIREDVNRNYYVADLTVEVVYTSQM-ALKWI	181
DB	121	DELLRVGVIEIYNEKIYDLL--NKKNQDLKIHESGNIVNVN--CBECIITSEVDLRL	176
QY	182	TKGEXSRHYGETYMNQORSRSHITFRMILESEKPEPSNCEGSVKYSHLNLYDLASERA	241
DB	177	CLGNKERTVGETYMNERSRSHAIFKIITESR-KSDHSD-DDAVIOSVLNVLDASERA	234
QY	242	AQTGAAGVPLKSGCININRSLFILGOVKKLSDQGVGFYINRDSKLTIRLQNSLGGNPKT	301
DB	235	DOTGARGALKSGGHINKLLFLSNVKSLENADNRNFYINRDSKLTIRLQASLGGNAFT	294
QY	302	RIICITPVSFDETTALQFASTAKYMKNTPTVYNEV	337
DB	295	SIICITKPSIMEESQSTLSFATRAKKIRIKPQVNEM	330
RESULT	11		
Q8RW4			
ID	Q8RW4	PRELIMINARY; PRT; 1055 AA.	
AC	Q8RW4,		
DT	01-JUN-2002 (TEMBLrel. 21, Created)		
DT	01-JUN-2002 (TEMBLrel. 21, Last sequence update)		
DT	01-OCT-2003 (TEMBLrel. 25, Last annotation update)		
DE	Putative kinesin.		
GN	AT4G39050.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.		
NCBI	TaxID=3702;		
EN	NCBI		
RX	[1]		
RP	SEQUENCE FROM N.A.		
RA	Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,		
RA	Deng J.M., Goldsmith A.D., Lee J.M., Oncdera C.S., Quach H.L.,		
RA	Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,		
RA	Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,		
RA	Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,		
RA	Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,		
RA	Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,		
RA	David R.W., Ecker J.R., Theologis A.;		
RA	"Arabidopsis Full Length cDNA Clones.;"		
RL	submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AY091060; AAM13881.1; -		
DR	GO; GO:0005871; C:kinesin complex; IEA.		
DR	GO; GO:0005524; F:ATP binding; IEA.		
DR	GO; GO:0003774; F:motor activity; IEA.		
DR	GO; GO:0007017; P:microtubule-based process; IEA.		
DR	InterPro; IPR001752; kinesin motor.		
DR	Pfam; PF00225; kinesin; I.		
DR	PRINTS; PS00380; KINESINHEAVY.		
DR	SMART; SM00129; KISC; 1.		
DR	SMART; SM00184; RING; 1.		
DR	PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.		
DR	PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.		
DR	PROSITE; PS50089; ZF RING 2; 1.		
DR	PROSITE; 1055 AA; _116452 MW; BBF3709ACB40215B CRC64;		

Query Match	39.3%;	Score 686;	DB 10;	Length 1055;
Best Local Similarity	46.2%;	Pred. No. 2.6e-45;		
Matches	159;	Conservative 57;	Mismatches 114;	Indels 14; Gaps 7;

Qy	1	A	E	G	A	V	C	V	R	P	L	N	S	R	E	S	L	G	E	T	A	Q	V	Y	K	T	N	N	V	I	Q	D	G	K	S	F	N	F	R	V	H	G	N	E	T	T	K	N	60						
Db	94	S	E	R	D	S	I	S	V	T	R	F	R	P	L	S	R	E	T	F	Q	R	G	E	V	A	N	Y	P	D	G	T	L	V	R	H	E	N	P	L	T	A	F	A	D	K	V	F	Q	P	A	T	I	D	153
Qy	61	V	Y	E	I	A	P	I	I	D	S	A	T	O	G	N	T	I	F	A	G	O	T	A	S	G	T	T	T	M	G	S	E	D	H	L	G	V	P	I	A	I	H	I	F	O	K	I	K	E	P	120			
Db	154	V	Y	D	V	A	P	P	V	K	A	M	E	G	V	N	G	T	V	F	A	G	V	T	S	S	G	K	T	H	T	M	H	G	D	Q	E	S	P	I	I	A	K	D	V	F	S	I	O	T	P	213			
Qy	121	D	R	E	F	L	R	V	S	Y	M	E	I	N	E	T	I	T	L	L	C	T	Q	K	K	P	L	I	E	D	V	N	R	N	V	Y	V	A	D	L	E	E	V	V	T	S	M	A	L	K	180				

```

Db 214 GREFLRVSYLRIYNEVINDLLDPTG--QNLRVRED--SQGTVEGKEEVLSPGHLSF 270
QY 181 IYKGEKSRHYGETKMNORSSRSHITFRMILRESKEGPEPNCESGKVKVSHNLVDLAGSER 240
Db 271 IAAGEHRHVGNFNLLSRSHITFTLMVESATGDEYD--GVTFSQLNLDLAGSE- 326
QY 241 AAOQGAAGVRLKEGNCNINSLFILGVIKKLDGQGVGFYINRDSKLTILQNSLGGNPK 300
Db 327 SSKTETGLRRKEGKSVYINKSLTLGTIVICKLSEK-ATHIPYRDSKLTLLQSSLSGHG 385
QY 301 TRIICTITPV--SPDETITAFQSTAK-----YMKNTPTVNEVS 338
Db 386 VSLICTITPASSSEETHNTLKFASRAKSIETIYASRNQIIDEKS 429

RESULT 12
QW5R5
ID QW5R5 PRELIMINARY; PRT; 1055 AA.
AC QW5R5;
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Kinesin-related protein (Putative kinesin protein).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_taxID=3702;
EN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21563048; PubMed=11706156;
RA Itoh R., Fujiwara M., Yoshida S.;
RT "Kinesin-Related Proteins with a Mitochondrial Targeting Signal.";
RL Plant Physiol. 127:724-726(2001).
EN [2]
RP SEQUENCE FROM N.A.
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
RA Quach H.L., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,
RA Palm C.J., Shinn P., Southwick A., Tripp M.G., Wu T., Davis R.W.,
RA Ecker J.R., Theologis A.;
RT "Arabidopsis Open Reading Frame (ORF) Clones.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB062739; BAB71852.1; -
DR EMBL; AY150516; AAN13032.1; -
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IPR01752; Kinesin_motor.
DR Pfam; PF00225; kinesin_1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN; 1.
DR PROSITE; PS50067; KINESIN MOTOR_DOMAIN2; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.
SQ SEQUENCE 1055 AA; 116463 MW; B69EC383FF14AB7B CRC64;

Query Match 39.3%; Score 686; DB 10; Length 1055;
Best Local Similarity 46.2%; Pred. No. 2,6e-45;
Matches 159; Conservative 57; Mismatches 114; Indels 14; Gaps 7;

QY 1 AEGNAVAVRVPLNRESLSGETAQVYWKTDNNVIYQVDSKSFNDFRVFHGNETTKN 60
Db 94 SERDSISVTVFRPLSDREYQRQGEVAVYDGTTLVRHEYNPLTAYAFDVKVFGPQATID 153
QY 61 VYEIEAIPIDSALQGYNGTIFAYGQTASCTYTMGSEDLGVIPRAIHDIQTKKFP 120
Db 154 VYDVARPVVKAAGVNGTVFAYGVTSSGKTHMGDQESPGIIPAIKDVFSIIQDTP 213

```

```

QY 121 DREFLLRVSYMEIYNETITDILCGTQKMKPLIREDVNRNYYVADLTETEVYVYSEMAK 180
Db 214 GREFLRVSYLRIYNEVINDLLDPTG--QNLRVRED--SQGTVEGKEEVLSPGHLSF 270
QY 181 IYKGEKSRHYGETKMNORSSRSHITFRMILRESKEGPEPNCESGKVKVSHNLVDLAGSER 240
Db 271 IAAGEHRHVGNFNLLSRSHITFTLMVESATGDEYD--GVTFSQLNLDLAGSE- 326
QY 241 AAOQGAAGVRLKEGNCNINSLFILGVIKKLDGQGVGFYINRDSKLTILQNSLGGNPK 300
Db 327 SSKTETGLRRKEGKSVYINKSLTLGTIVICKLSEK-ATHIPYRDSKLTLLQSSLSGHG 385
QY 301 TRIICTITPV--SPDETITAFQSTAK-----YMKNTPTVNEVS 338
Db 386 VSLICTITPASSSEETHNTLKFASRAKSIETIYASRNQIIDEKS 429

RESULT 13
QW5P8
ID QW5P8 PRELIMINARY; PRT; 699 AA.
AC QW5P8;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE OSM-3 protein (corresponding sequence M02B7.3b).
OS M02B7.3 OR OSM-3.
OC Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_taxID=6239;
EN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
EN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Bristol N2;
RA Nelson J., Wohldmann P.;
RT "The sequence of C. elegans cosmid M02B7.";
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
EN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; U70853; AAM97997.1; -.
DR Wormpep; M02B7.3b; CE31568.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin_1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN; 1.
DR PROSITE; PS50067; KINESIN MOTOR_DOMAIN2; 1.
SQ SEQUENCE 699 AA; 78779 MW; 8A774E3EF3A07813 CRC64;

Query Match 39.0%; Score 680; DB 5; Length 699;
Best Local Similarity 46.4%; Pred. No. 4.2e-45;
Matches 159; Conservative 53; Mismatches 109; Indels 22; Gaps 9;

QY 5 AVAVCVRVPLNRESLSGETAQVYWKTDNNVIYQV-----DG-SKSFNDFRVFHGNETT 58
Db 4 SVRVAVRCRPFNQREKDLNTTLCV-GMTFN--VGQVNLNAPDGAAXDFTFDGAYFMDSTG 60
QY 59 KNVYIEAIPIDSALQGYNGTIFAYGQTASCTYTMGSED---HLGVIPRAIHDIQK 115

```

GO; GO:0007017; P-microtubule-based process; IEA.

DR InterPro; IPR001752; kinesin_motor.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
DR PROSITE; PS0089; ZF_RING_2; 1.
DR PROSITE; PS0089; ZF_RING_2; 1.
SEQUENCE 1058 AA; 117089 MW; F97660F723A9A506 CRC64;

Query Match 38.6%; Score 674; DB 10; Length 1058;
Best Local Similarity 45.1%; Pred. No. 2.3e-44;
Matches 153; Conservative 61; Mismatches 115; Indels 10; Gaps 6;

QY 1 AERGAVAVCVRVPLNSREESLGETAQVYWKYTDNNVYIQVDGSKSFNDFRVFHNQETKN 60
DB 100 SERDSISVTVRFPMSEREYQRGDEIVWYDPADKMVRNEXNPLTAYAFDKVFGPQSTTPE 159
QY 61 VYEIETAPIIDSALQYNGTIFAYGQTASCKTYTMGSEDLGVIPRAIHDFIKIKKFP 120
DB 160 VYDVAAPVYKAAMEGVNGTVFAYGVTSCKTHTMGDQDPGPIIPIAKVDVFIIOETT 219
QY 121 DRFELRVSYWEIYNETITDLCTQKMKPLIIREVDNRNYYVADLTTEEYVYVYSEMALKW 180
DB 220 GREFLRVSVLEIYNEVINDLDDPTG--QNLRIRED--SGTYVEGIEKEEVLSFGHALSF 276
QY 181 ITGKESRHYGETKMNQRRSSRSHITFRMILESREKGEPSNCEGSKVKVSHLNLVDLAGSER 240
DB 277 IAGEEHRHVGSNNFNLMSRSKSHITFTLMTIESSAHGDQYD---GVIFSQLNLIIDLAGE- 332
QY 241 AAGTGAAGVRLKEGCNINRSFLTGQVTKKLSGQGVGFNVYRDSKLTIRLQNSILGNGPK 300
DB 333 SSKTETTGRLRKEGAYINKSLTGTGIGKUTGECT--THVFPKSLTRLLQSSLSGHGH 391
QY 301 TRILCTITPV--SFDETLTALQFASAKYMKNTPTVYVNEV 337
DB 392 VSLICTVTPASSSTEETHNTLKFPASRAKRIENASRNKI 430

RESULT 15

Q9LHL9 PRELIMINARY; PRT; 1033 AA.

AC Q9LHL9;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Kinesin (Centromere protein) like heavy chain-like protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety PL,
RT TAC and BAC clones";
RL DNA Res. 7:217-221(2000).
DR EMBL; AF002040; BAB03114.1; --
DR HSSP; P33176; 1BG2.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:26:01 ; Search time 24.3176 Seconds
(without alignments)
4694.096 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_405
Perfect score: 2053
Sequence: 1 AEEGAVVCVRPLNSREE.....IQKVQNEKIENLRMLVTSS 404

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2053	100.0	2633	4	ABG06505 Novel hum
2	2053	100.0	2663	4	Aam39097 Human pol
3	1944.5	94.7	2688	4	Aam40883 Human pol
4	1468.5	71.5	2954	2	Aay01632 Amino aci
5	820	39.9	366	4	Abu53125 Intracell
6	781	38.0	348	4	Abu53208 Human cel
7	754	36.7	2013	4	Abb62322 Drosophil
8	704	34.3	677	4	Abb65183 Drosophil
9	694.5	33.8	473	5	Abg70992 Human tar
10	694.5	33.8	522	5	Abg70991 Human Hsk
11	694.5	33.8	1232	7	Ad44938 Human lun
12	694.5	33.8	1232	7	Ad44938 Human lun
13	694	33.8	1034	3	Ad44938 Human lun
14	694	33.8	1069	3	Ad44938 Human lun
15	694	33.8	1121	3	Ad44938 Human lun
16	691.5	33.7	1029	7	Ad44938 Human lun
17	691.5	33.7	1029	7	Ad44938 Human lun
18	684.5	33.3	1232	7	Ad44938 Human lun
19	684.5	33.3	1232	7	Ad44938 Human lun
20	683	33.3	1038	5	Aam48337 Human kin
21	682.5	33.2	1234	5	Abg70993 Human Hsk
22	678.5	33.0	1232	5	Abg70993 Human Hsk
23	662.5	32.3	1048	4	Abg70990 Human Hsk
24	662.5	32.3	1066	4	Abg70990 Human Hsk
25	662.5	32.3	1066	6	Abg72693 Fruitfly

26	652	31.8	411	2	AAW72745	Drosophil
27	652	31.8	441	2	AAW72744	Drosophil
28	652	31.8	975	2	AAW72746	Drosophil
29	652	31.8	975	4	ABB63485	Drosophil
30	652	31.8	975	7	ADB67088	Kinesin h
31	652	31.8	1388	5	AAE14400	Human kin
32	652	31.8	1388	5	AAU79590	Human kin
33	652	31.8	1388	6	ABR48222	Human kin
34	652	31.8	1388	7	ADB80468	Ovarian c
35	652	31.8	1388	7	ADC35116	Human bre
36	651.5	31.7	757	4	AAU19569	Human dia
37	651.5	31.7	757	5	ABP51294	Human MDD
38	649	31.6	796	7	ADE55349	Rat Prote
39	649	31.6	1921	4	ABE62362	Drosophil
40	647.5	31.5	829	3	AAU79590	Human kin
41	647.5	31.5	834	3	AAU79590	Human kin
42	645	31.4	1031	7	ABG67093	Kinesin h
43	644.5	31.4	762	5	ABG60124	Human DIT
44	644.5	31.4	1362	5	AAU74840	Human Hsk
45	644.5	31.4	1805	5	ABP68930	Human pol

ALIGNMENTS

RESULT 1
ABG06505
ID ABG06505 standard; protein; 2633 AA.

XX AC ABG06505;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #6496.

XX KW Human; chromosome mapping; Gene mapping; Gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Dmanac RT, Liu C, Tang YT;

XX XX WPI; 2001-639362/73.

XX DR N-PSDB; AAS70692.

XX PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

XX PS Claim 20; SEQ ID NO 36864; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying normal genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging

CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2633 AA;

Query Match 100.0%; Score 2053; DB 4; Length 2633;
Best Local Similarity 100.0%; Pred. No. 1.5e-171;
Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AEEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDFRVEHNETTKN 60
DB 2 AEEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDFRVEHNETTKN 61
QY 61 VYEIAAIIIDSAIQGVNGTIFAYGOTASGKTYTMGSEDLGVIPRAIHDFQKIKKPP 120
DB 62 VYEIAAIIIDSAIQGVNGTIFAYGOTASGKTYTMGSEDLGVIPRAIHDFQKIKKPP 121
QY 121 DREFLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVYVADLVEEVVYSEMAKWK 180
DB 122 DREFLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVYVADLVEEVVYSEMAKWK 181
QY 181 ITKGEKSRHYGETKMNORSRSHTIFRMILESREKGEPSNCEGSKVSHNLVLDLAGSER 240
DB 182 ITKGEKSRHYGETKMNORSRSHTIFRMILESREKGEPSNCEGSKVSHNLVLDLAGSER 241
QY 241 AAQTGAAGVRLKEGCNINRSLFILGVQVKKLSQGVGGFINYRDSKLTILQNSLGNPK 300
DB 242 AAQTGAAGVRLKEGCNINRSLFILGVQVKKLSQGVGGFINYRDSKLTILQNSLGNPK 301
QY 301 TRIICTITPVSFDETLTALQFASAKYMKNTPYNEVSTDEALLKRYRKEIMDLKKOLEE 360
DB 302 TRIICTITPVSFDETLTALQFASAKYMKNTPYNEVSTDEALLKRYRKEIMDLKKOLEE 361
QY 361 VSLTRAQAMEKQDLAQLLEEKDLQKQVNEKIENLTRLMTSS 404
DB 362 VSLTRAQAMEKQDLAQLLEEKDLQKQVNEKIENLTRLMTSS 405

RESULT 2
AAM39097
ID AAM39097 standard; protein; 2663 AA.
XX
AC AAM39097;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2242.
XX
KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
XX 23-DEC-1999; 99US-00471275.
PR

PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
XX Zhou P, Goodrich R, Drmanac RT;
DR WPI; 2001-442253/47.
DR N-PSDB; AAI58253.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
XX as central nervous system injuries.
PT
PT
XX Example 4; SEQ ID NO 2242; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
XX encoded polypeptides (AAM38642-AAM42213) with neotropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders. Note: The sequence data for this patent did not form
XX part of the printed specification
XX
SQ Sequence 2663 AA;

Query Match 100.0%; Score 2053; DB 4; Length 2663;
Best Local Similarity 100.0%; Pred. No. 1.5e-171;
Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AEEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDFRVEHNETTKN 60
DB 2 AEEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDFRVEHNETTKN 61
QY 61 VYEIAAIIIDSAIQGVNGTIFAYGOTASGKTYTMGSEDLGVIPRAIHDFQKIKKPP 120
DB 62 VYEIAAIIIDSAIQGVNGTIFAYGOTASGKTYTMGSEDLGVIPRAIHDFQKIKKPP 121
QY 121 DREFLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVYVADLVEEVVYSEMAKWK 180
DB 122 DREFLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVYVADLVEEVVYSEMAKWK 181
QY 181 ITKGEKSRHYGETKMNORSRSHTIFRMILESREKGEPSNCEGSKVSHNLVLDLAGSER 240
DB 182 ITKGEKSRHYGETKMNORSRSHTIFRMILESREKGEPSNCEGSKVSHNLVLDLAGSER 241
QY 241 AAQTGAAGVRLKEGCNINRSLFILGVQVKKLSQGVGGFINYRDSKLTILQNSLGNPK 300
DB 242 AAQTGAAGVRLKEGCNINRSLFILGVQVKKLSQGVGGFINYRDSKLTILQNSLGNPK 301
QY 301 TRIICTITPVSFDETLTALQFASAKYMKNTPYNEVSTDEALLKRYRKEIMDLKKOLEE 360
DB 302 TRIICTITPVSFDETLTALQFASAKYMKNTPYNEVSTDEALLKRYRKEIMDLKKOLEE 361
QY 361 VSLTRAQAMEKQDLAQLLEEKDLQKQVNEKIENLTRLMTSS 404
DB 362 VSLTRAQAMEKQDLAQLLEEKDLQKQVNEKIENLTRLMTSS 405

RESULT 3
AAW40883
ID AAW40883 standard; protein; 2688 AA.
XX
AC AAW40883;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 5814.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
FN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-0052317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
WPI: 2001-442253/47.
DR N-PSDB; AAI60039.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
PS Example 2; SEQ ID NO 5814; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAW38642-AAW42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 2688 AA;
XX

Query Match 94.7%; Score 1944.5; DB 4; Length 2688;
Best Local Similarity 96.1%; Pred. No. 6e-162;
Matches 391; Conservative 2; Mismatches 11; Indels 3; Gaps 3;
XX 1 ABEAGAVACVVRPLNSRRESLGETAQVYWKTHNNVIYQVDSKSFNDRVHNETTKN 60

Db 23 ABEAGAVACVVRPLNSRRESLGETAQVYWKTHNNVIYQVDSKSFNDRVHNETTKN 82
QY 61 VYEETAAPIIDSAIQYNGTIFAYGQTASGKTYTMMGSDHLGVIPRA-IHDFP-QKIKK 118
Db 83 VYEETAAPIIDSAIQYNGTIFAYGQTASGKTYTMMGSDHLGVIPQGPHGHFSKIXE 142
QY 119 -FPDRFLLRVSYMEIYNETITDILCGTQKMKPLIIREDNRRNVYVADLTEEYVYSEMA 177
Db 143 VFDRFLLRVSYMEIYNETITDILCGTQKMKPLIIREDNRRNVYVADLTEEYVYSEMA 202
QY 178 LKWITKGEKSRHYGETKMQRSRSHITFRMILESEKGEPCNCEGSKVSHNLVDLAG 237
Db 203 LKWITKGEKSRHYGETKMQRSRSHITFRMILESEKGEPCNCEGSKVSHNLVDLAG 262
QY 238 SRAAQTGAAGVRLKEGCNINRSLFILGOVKKLSGQGVGFNRYRDSKLTILQNSLGG 297
Db 263 SRAAQTGAAGVRLKEGCNINRSLFILGOVKKLSGQGVGFNRYRDSKLTILQNSLGG 322
QY 298 NPKTRIICTITPVSFDETLTALQFASTAKYMQNTPYVNEVSTDEALLKRYRKEIMDLKKQ 357
Db 323 NPKTRIICTITPVSFDETLTALQFASTAKYMQNTPYVNEVSTDEALLKRYRKEIMDLKKQ 382
QY 358 LEEVSLTETRAQAMEKDQLAQLLEEKDLLQKVQNEKIENLTETMTSS 404
Db 383 LEEVSLTETRAQAMEKDQLAQLLEEKDLLQKVQNEKIENLTETMTSS 429

RESULT 4
AAW01632
ID AAW01632 standard; protein; 2954 AA.
XX
AC AAW01632;
XX
DT 22-JUN-1999 (first entry)
XX
DE Amino acid sequence of centromere-associated protein-E (CENP-E).
XX
KW CENP-E; centromere-associated protein-E; ATPase activity;
KW plus end-directed microtubule motor activity; chromosome congression;
KW microtubule binding activity; chromosome movement; mitosis;
KW cell proliferation; tumor; metastasis; vascular malfunction;
KW inflammatory disease; immune disease; angiogenesis; hypertension;
KW restenosis; fungal infection; selective herbicide; fungicide;
KW insecticide; plant growth regulator; activator; cancer cell marker.
XX
OS Xenopus sp.
XX
FN WO9913061-A1.
XX
PD 18-MAR-1999.
XX
PF 10-SEP-1998; 98WO-US019231.
XX
PR 11-SEP-1997; 97US-0058645P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Wood KW, Sakowicz R, Goldstein LSB, Cleveland DW;
XX
DR WPI; 1999-229233/19.
XX
DR N-PSDB; AAW26819.
XX
PT Centromere-associated protein-E and related nucleic acid.
XX
PS Claim 5; Page 66-67; 77pp; English.
XX
CC The present sequence represents CENP-E (centromere-associated protein-E)
CC of Xenopus. The protein has at least one of plus end-directed microtubule
CC motor activity, ATPase (adenosine triphosphatase) activity and
CC microtubule binding activity. CENP-E is the motor that powers chromosome
CC movement toward microtubule plus ends and is essential for congression of
CC chromosomes during mitosis. Modulators of CENP-E can thus control cell

XX	(GEHU-) GERMAN HUMAN GENOME PROJECT.
XX	WIEMANN S;
XX	WPI; 2001-327840/34.
XX	Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies.
PS	Example III; Page 829; 1095pp; English.
XX	This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence represents a homologue of a polypeptide described in the disclosure of the invention
XX	Sequence 348 AA;
Query Match	38.0%; Score 781; DB 4; Length 348;
Best Local Similarity	51.1%; Pred. No. 2.6e-60;
Matches 178; Conservative 38; Mismatches 106; Indels 26; Gaps 7;	
QY	11 RVRPLNSREELGETAQYW-----KTDNNVIQVDGSGFNDFRVFHGNETTKNVYEE 64
Db	1 RCRPLNEREINDGCSCVQVPPTGYKTIVNHGEGDSPHKSFTHDVFWMCTQEDVVDT 60
QY	65 IAAPIIDSAIOGYNGTIFAYCQTASGKTYTMWG-----SEDLGVIPRAIHDIPOKIKKFP 120
Db	61 VAHPDVDCEFGYNCTTFAYCQTGSKTYTMWPGGEPDHMGILPRCCDHIDRIDKFQ 120
QY	121 DRE---FLLRVSVMIEIYNETITDLLC-GTOKMKPLINREDVNRYVADLTTEEVYTSEM 176
Db	121 EKDHDFHWKCSVMIEIYNEEYDILLCPNPQMKNLNIHEHNMGPYVGCTEFHVCSVED 180
QY	177 ALKWITKGESRRHYGETKMQRSSRSHTIPRMILESRKEGPSNCGSVKYSHLINIYDLA 236
Db	181 ACHWIWGQNKNRHVAATMNDHSRSHTIFTIHVEQRHK-----QCDEHVCBSKMNLYDLA 236
QY	237 GSERAQTGAAGVLKKGNCINRSLFILGOVKKLSDGQV-----GGFINYRDSKLTRI 290
Db	237 GSERVNRGTAEQRLKEGNCINQLTLGNVINALADQTKMYGGHGHLFYRDSKLTLWL 296
QY	291 LQNSLGNPKTRICTITTPV--SFDFTILTALQFASTAKYMKNTPYNE 336
Db	297 LQDSLGNCKTCMIACIWPADWNYYEETLSTIRYADRANKIKKPQINE 344
RESULT 7	
ID	ABB62322
ID	ABB62322 standard; protein; 2013 AA.
AC	ABB62322;
XX	
DT	26-MAR-2002 (first entry)
DE	Drosophila melanogaster polypeptide SEQ ID NO 13758.
XX	Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
KW	
XW	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	

Db 66 AVAPLIKGVFGYNATVLAQTGSGKTSYMGAYTAQENEPVGVIPRVQLLFXEID 125
 Qy 118 KFPDREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVYVADLTTEVVYTSEMA 177
 Db 126 KKSDFEFLKVSYLEIYNEEILDLCPREKAQINREDPKGKIVGLTEKTVLVALDT 185
 Qy 178 LKWITKGEKSRHYGETKMNORSRSHTIFRMILRESREKGPSCGSKVSHLNLVDLAG 237
 Db 186 VSCLEQGNNSRTVASTAMNSQSSRSHAFITISLEQRKSD----KNSFRSKLHVLVDLAG 241
 Qy 238 SERAAQTGAAGVRLKEGCNINRSILFGLQVVKILSDGQVGFNYRDSKLTILQNSLGG 297
 Db 242 SERQKTKAEGDRLEKGININRGLCLGNVISALGDDKGGFVPRYRDSKLTLLQDSLGG 301
 Qy 298 NPKTRIICTITPV--SFDETLTALQFASAKYMKNTPVYNEVSTDEALLKRYKEIMDLK 355
 Db 302 NSHTLMACVSPADSNLEETILTRYADRARKINKPIVN-----IDPQTAEHLNHLK 353
 Qy 356 KQLEEVSL-----ETRAQAMEKDQLAQLLEEKDILQKVNQKIENTL 396
 Db 354 QVQQLQVLLLOAHGGTLPGSITVPESENLSLMEKNQ--SLVEENEKLSRGLSEAAQT 411
 Qy 397 TRML 400
 Db 412 AQML 415

RESULT 10
 ABG70991

ID ABG70991 standard; protein; 522 AA.

AC ABG70991;

DT 10-DEC-2002 (first entry)

DE Human Hskif4 construct protein.

XX Microtubule stimulated ATPase; cellular proliferation; cancer; AIDS;
 KW hyperplasia; restenosis; cardiac hypertrophy; immune disorder; HIV;
 KW inflammation; autoimmune disease; arthritis; graft rejection; psoriasis;
 KW inflammatory bowel disease; proliferation; medical procedure; surgery;
 KW human immunodeficiency virus; acquired immunodeficiency syndrome;
 KW angioplasty; human; Hskif4; kinesin family.

OS Homo sapiens.

Key Location/Qualifiers

FT Region 1..22

FT /note= "N-terminal T7 epitope"

FT Region 496..516

FT /note= "C-terminal myc epitope"

FT Region 517..522

FT /note= "6-histidine residues at C-terminus"

XX US6440684-B1.

XX 27-AUG-2002.

XX 12-JUN-2000; 2000US-00592054.

XX 12-JUN-2000; 2000US-00592054.

XX (CYTO-) CYTOKINETICS INC.

XX Braud C, Finer JT, Sakowicz R, Wood KW;

XX WPI; 2002-711529/77.

XX N-PSDB; ABS55161.

XX Screening for modulators of target protein having microtubule stimulated
 PT ATPase activity e.g. kinesin family of protein, useful for treating
 PT cancer, psoriasis, arthritis, human immunodeficiency virus (HIV)

PT infection.

PS Claim 2; Fig 4; 34pp; English.

XX The present invention relates to a new method of screening modulators of
 CC target protein with microtubule stimulated ATPase activity. The method
 CC involves contacting the target protein with an agent at 1st and 2nd
 CC concentrations and determining the level of activity (e.g. binding or
 CC ATPase activity) of target protein, where a difference between levels of
 CC activity of target protein contacted with 1st and 2nd concentrations of
 CC an agent indicates that an agent modulates activity of target protein.
 CC The invention can be used for screening for modulators of target protein
 CC having microtubule stimulated ATPase activity. The compounds identified
 CC by method of the invention are useful for treating cellular proliferation
 CC including cancer, hyperplasias, restenosis, cardiac hypertrophy, immune
 CC disorders and inflammation. The compounds identified by the method are
 CC also useful for treating autoimmune disease, arthritis, graft rejection,
 CC inflammatory bowel disease, proliferation induced by medical procedures,
 CC e.g. surgery, angioplasty etc. The compounds are also useful for treating
 CC psoriasis. The compounds are useful for inhibiting human immunodeficiency
 CC virus (HIV) and thus treating acquired immunodeficiency syndrome (AIDS).
 CC The present amino acid sequence represents the human Hskif4 (kinesin
 CC family) construct protein of the invention

XX Sequence 522 AA;

Query Match 33.8%; Score 694.5; DB 5; Length 522;

Best Local Similarity 39.6%; Pred. No. 2.1e-52;

Matches 168; Conservative 66; Mismatches 145; Indels 45; Gaps 8;

Qy 6 VAVCVRVPLNSREESLG-ETAQVYMTDNNVIYQVDSKSFNDFRVFHGNETTKNVVEE 64

Db 30 VRVALRCPLVPKKEISGCGMCLSFVPEGEQVVGTD--KSFTYDFVDPSTGEVEVNT 87

Qy 65 IAAPIDSAIQXNGTIFPAYQGTASGKTYTMG-----SEHLGVIPRAIHDFQIK 117

Db 88 AVAPLIKGVFGYNATVLAQTGSGKTSYMGAYTAQENEPVGVIPRVQLLFXEID 147

Qy 118 KFPDREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVYVADLTTEVVYTSEMA 177

Db 148 KKSDFEFLKVSYLEIYNEEILDLCPREKAQINREDPKGKIVGLTEKTVLVALDT 207

Qy 178 LKWITKGEKSRHYGETKMNORSRSHTIFRMILRESREKGPSCGSKVSHLNLVDLAG 237

Db 208 VSCLEQGNNSRTVASTAMNSQSSRSHAFITISLEQRKSD----KNSFRSKLHVLVDLAG 263

Qy 238 SERAAQTGAAGVRLKEGCNINRSILFGLQVVKILSDGQVGFNYRDSKLTILQNSLGG 297

Db 264 SERQKTKAEGDRLEKGININRGLCLGNVISALGDDKGGFVPRYRDSKLTLLQDSLGG 323

Qy 298 NPKTRIICTITPV--SFDETLTALQFASAKYMKNTPVYNEVSTDEALLKRYKEIMDLK 355

Db 324 NSHTLMACVSPADSNLEETILTRYADRARKINKPIVN-----IDPQTAEHLNHLK 375

Qy 356 KQLEEVSL-----ETRAQAMEKDQLAQLLEEKDILQKVNQKIENTL 396

Db 376 QVQQLQVLLLOAHGGTLPGSITVPESENLSLMEKNQ--SLVEENEKLSRGLSEAAQT 433

Qy 397 TRML 400

Db 434 AQML 437

RESULT 11

ADD49938

ID ADD49938 standard; protein; 1232 AA.

XX ADD49938;

XX 15-JAN-2004 (first entry)

XX Human lung specific tumour antigen L 1447p.

KW Human; lung cancer antigen; cytostatic; lung cancer; gene therapy;
KM vaccine; T-cell; tumour.
XX
OS Homo sapiens.
XX
FN US2003194764-A1.
XX
PD 16-OCT-2003.
XX
PF 04-APR-2002; 2002US-00116712.
XX
XX 05-APR-2001; 2001US-0282289P.
PR 05-OCT-2001; 2001US-0327511P.
XX
XX (CORI-) CORIXA CORP.
PA
XX Bangur CS, Switzer A;
XX WPI; 2003-844452/78.
XX N-PSDB; ADD49936.
DR
DR New isolated polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cancer, particularly lung cancer.
PT
XX
XX Claim 1; SEQ ID NO 670; 250pp; English.
XX
XX The invention relates to an isolated polynucleotide (a) comprising any of
CC the 666 fully defined nucleotide sequences appearing as ADD49269 -
CC ADD49936, ADD49936 and ADD49938, complements of (a); sequences of (a)
CC at least 20 contiguous residues of (a); sequences that hybridise to (a)
CC under highly stringent conditions; sequences having at least 75 or 90%
CC identity to (a); or degenerate variants of (a). Also included are an
CC isolated polypeptide (b) (comprising: sequences encoded by the new
CC polynucleotide; any of the 4 amino acid sequences fully defined in the
CC polynucleotide; or sequences having at least 70 or 90% identity to the
CC sequence in (a) or (b)), an expression vector comprising the above, a host
CC polynucleotide operably linked to an expression control sequence, a host
CC cell transformed or transfected with the above expression vector, an
CC isolated antibody, or its antigen-binding fragment, that specifically
CC binds to the above polypeptide, an oligonucleotide that hybridises to the
CC above-mentioned nucleotide sequences under highly stringent conditions, a
CC fusion protein comprising at least one polypeptide cited above, detecting
CC the presence of a cancer in a patient (comprising: obtaining a biological
CC sample from the patient; contacting the biological sample with a binding
CC agent that binds to the polypeptide, or with the oligonucleotide cited
CC above; detecting in the sample an amount of the polypeptide that binds to
CC the binding agent, or an amount of a polynucleotide that hybridises to
CC the oligonucleotide; and comparing the amount of polypeptide, or
CC polynucleotide that hybridises to the oligonucleotide, to a predetermined
CC cut-off value and then determining the presence of a cancer in the
CC patient), a method for stimulating and/or expanding T-cells specific for
CC a tumour protein (comprising contacting T-cells with the above
CC polypeptide, polynucleotide or antigen-presenting cells that express the
CC polynucleotide, under conditions and for a time sufficient to permit the
CC stimulation and/or expansion of T-cells), an isolated T-cell population
CC comprising T-cells prepared by the method, a composition comprising a
CC first component selected from physiological carriers and
CC immunostimulants, and a second component selected from the above
CC polypeptide, polynucleotide, antibody, fusion protein, T-cell population
CC and antigen-presenting cells that express the above polypeptide, to
CC stimulating an immune response in a patient (comprising administering to
CC the patient the above composition) treating lung cancer in a patient
CC (comprising administering to the patient the above composition and a
CC diagnostic kit (comprising: at least one oligonucleotide cited above; or
CC at least one antibody cited above and a detection reagent, where the
CC detection reagent comprises a reporter group). The composition and
CC methods are useful in diagnosing, preventing and treating cancer,
CC particularly lung cancer. The present sequence is a lung cancer-
CC associated antigen of the invention.

SQ Sequence 1232 AA;

Query Match 33.8%; Score 694.5; DB 7; Length 1232;

Best Local Similarity 39.6%; Pred. No. 7.5e-52;
Matches 168; Conservative 66; Mismatches 145; Indels 45; Gaps 8;
QY 6 VAVCVVRPLMSREESLG-ETAQVYWKTDNNVYVQDGSKGFNDRVPHGNETTTNNVVE 64
DB 10 VRVALRCPLVPKEISEGQMCLSFVPGEPQVVVGTD--KSTYDFVDFPSTEQEVEVENT 67
QY 65 IAAPIDSIAOQYNGTIFAYGQTASGKYTYMG-----SEDHLGVIPRAHDIFQKIK 117
DB 68 AVAPLIKGVFGYNATVLAAYGQTSGKTYSGMGAYTAQENEPYGVIPRVQLLFKEID 127
QY 118 KPDEFEFLRYSYMEIYNETIDLLCGTQKMKPLIIREDVNRNVVADLTESVWVTSEMA 177
DB 128 KKSDFEFTLKVSLEYIENEILLDLCPSREKQAINREDPKGKIIVGLTEKTVLVALDT 187
QY 178 LKWITKGEKSRHYGETKMNQSRSHITFRMILESREKGEPCNCGSVKSHNLNVDLAG 237
DB 198 VSCLEQGNNSRTVASTAMNSQSSRSHAFTTISLEQRKKS-----KNSSPRSKLHVDLAG 243
QY 238 SERAAQTGAAGVRLKEGNCNINRSFILGQVINKLSDGQVGGFNYRDSKLTILQNSLGG 297
DB 244 SERQKTKAEGDRLKEGININRGLLCLGNVISALGDDKGGFVYRDSKLTILQNSLGG 303
QY 298 NPKTRILICTITPV--SEDETLTALOFASTAKYMKNTPYVNEVSTDEALKRYRKEIMDK 355
DB 304 NHTLMIACVSPADNLEETNLTRYADPARKIKKPIV-----IDPQTAEHLNHLK 355
QY 356 KOLEEVSL-----ETRAQAMKQQLAQLEEKDLLQKVONEKIENL 396
DB 356 QOVQQLQVLLQAHGQTLPGSITVPESENLSLMEKNQ--SLVEENEKLSRGLSEAAQGT 413
QY 397 TRML 400
DB 414 AQML 417

RESULT 12

ADD18924
ID ADD18924 standard; protein; 1232 AA.

XX AC ADD18924;

XX DT 15-JAN-2004 (first entry)

XX DE Human disease related protein SeqID413.

XX KW human; disease state; cytostatic; antiinflammatory; ophthalmological;
KW antiarteriosclerotic; vulnary; gene therapy;
KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;
KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;
KW glucose transport; catecholamine synthesis; iron transport;
KW nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;
KW retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;
KW inflammatory condition; wound healing.

XX OS Homo sapiens.

XX PN WC2003018621-A2.

XX PD 06-MAR-2003.

XX PF 23-AUG-2002; 2002WO-GB003892.

XX PR 23-AUG-2001; 2001GB-00020558.

XX PR 05-OCT-2001; 2001GB-00024037.

XX PA (OXFO-) OXFORD BIOMEDICA UK LTD.

XX PI Kingeman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;

XX DR WPI; 2003-290046/28.

XX DR N-PSDB; ADD18925.

PT New substantially purified polypeptide, useful for diagnosing or treating
 PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion
 PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or
 PT wound healing.

XX Claim 25; SEQ ID NO 413; 424pp; English.

XX This invention relates to novel human genes and gene product which are
 CC implicated in certain disease states. Compounds which modulate the
 CC proteins of the invention may have cytostatic, antiinflammatory,
 CC ophthalmological, antiarteriosclerotic or vulnerary activities. The
 CC sequences of the invention may be useful for gene therapy. The invention
 CC may be useful for diagnosing or treating a hypoxia-regulated condition,
 CC such as tumorigenesis, angiogenesis, apoptosis, inflammation,
 CC erythropoiesis, or the biological response to hypoxia conditions
 CC including processes such as glycolysis, gluconeogenesis, glucose
 CC transportation, catecholamine synthesis, iron transport or nitric oxide
 CC synthesis. The disease includes cancer, ischaemic conditions, reperfusion
 CC injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,
 CC inflammatory conditions or wound healing. The present sequence is that of
 CC a disease related protein of the invention.

XX SQ Sequence 1232 AA;

Query Match 33.8%; Score 694.5; DB 7; Length 1232;

Best Local Similarity 39.6%; Pred. No. 7.5e-52;

Matches 168; Conservative 66; Mismatches 145; Indels 45; Gaps 8;

QY 6 VAVCVVRPLNRSRLG-ETAAQVYKTDNNVIVQDGSFNPDRVPHGNETTKNYYEE 64

Db 10 VRVALRCEPLVPKEISEGCQCLSFVPGEPQVVGTGDK--KSFYDFVDPSTEQEEVENT 67

QY 65 IAAPIIDSAIQYNGTIPAYQGTASGKTYTMMG-----SEDLGVIPRAIHDFOKIK 117

Db 68 AVAPLKGVFGYNATLAYGTGSGKTYSMGAYTAQENEPVGVIPVLIQLFREID 127

QY 118 KFPDFEFLRYSMEIYNETTDLCTGQKWKPLIREDVNRNVYADLTVEEVYTSMA 177

Db 128 KKSDFEFLKVSYLEIYNEEILDLCPREKAQINREDPKGKIVGLTEKTVLVALDT 187

QY 178 LKWTGKESRHYGETKWNQSSSHTIFRMILSREKGEPSNCEGSKVSHMLNVLDA 237

Db 188 VSCLEQGNSTVASTANSSSHAFITLSLQKSD----KNSFSRSLHLVLDLAG 243

QY 238 SERAAQTGAAGVRLEKGCNINRSLFLLQVTKLSGQGVGFVNRVDSKLTILQNSLGG 297

Db 244 SEROKTKAEGDRLEKGINRGLGLGNVLSALGDDKGGFVYRSLKTRLLQDSLGG 303

QY 298 NPKTRIICITPV--SFDEITLQFASAKYMNQVYNEVSTDEALLKRYRKEIMDLK 355

Db 304 NSHTLMACVSPADSNLEETLTLRYADRAKIKNKEIVN-----IDPQTAELNHLK 355

QY 356 KOLEEVSLS-----ETRAQMEKDLQALQLEEKDLQKQVNEKLENL 396

Db 356 QVQVQLQVLLQAHGGLTPGSITVPSSENLQSLMEKNQ--SLVEENKLSRGLSEAAQGT 413

QY 397 TRML 400

Db 414 AQML 417

RESULT 13

AAG31112

ID AAG31112 standard; protein; 1034 AA.

XX AC AAG31112;

XX DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 37309.

DE Protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW

KW termination sequence.
 XX Arabidopsis thaliana.
 OS EP1033405-A2.
 PN 06-SEP-2000.
 PD 25-FEB-2000; 2000EP-00301439.
 PF 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.
 PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126264P.
 PR 29-MAR-1999; 99US-0126785P.
 PR 01-APR-1999; 99US-0127462P.
 PR 06-APR-1999; 99US-0128234P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0129845P.
 PR 19-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130449P.
 PR 23-APR-1999; 99US-0130510P.
 PR 23-APR-1999; 99US-0130891P.
 PR 28-APR-1999; 99US-0131449P.
 PR 30-APR-1999; 99US-0132048P.
 PR 30-APR-1999; 99US-0132407P.
 PR 04-MAY-1999; 99US-0132484P.
 PR 05-MAY-1999; 99US-0132485P.
 PR 06-MAY-1999; 99US-0132486P.
 PR 06-MAY-1999; 99US-0132487P.
 PR 07-MAY-1999; 99US-0132863P.
 PR 11-MAY-1999; 99US-0134256P.
 PR 14-MAY-1999; 99US-0134218P.
 PR 14-MAY-1999; 99US-0134219P.
 PR 14-MAY-1999; 99US-0134221P.
 PR 14-MAY-1999; 99US-0134370P.
 PR 18-MAY-1999; 99US-0134768P.
 PR 19-MAY-1999; 99US-0134941P.
 PR 20-MAY-1999; 99US-0135124P.
 PR 21-MAY-1999; 99US-0135353P.
 PR 24-MAY-1999; 99US-0135629P.
 PR 25-MAY-1999; 99US-0136021P.
 PR 27-MAY-1999; 99US-0136392P.
 PR 28-MAY-1999; 99US-0136782P.
 PR 01-JUN-1999; 99US-0137222P.
 PR 03-JUN-1999; 99US-0137528P.
 PR 04-JUN-1999; 99US-0137502P.
 PR 07-JUN-1999; 99US-0137724P.
 PR 08-JUN-1999; 99US-0138094P.
 PR 10-JUN-1999; 99US-0138540P.
 PR 10-JUN-1999; 99US-0138847P.
 PR 14-JUN-1999; 99US-0139119P.
 PR 16-JUN-1999; 99US-0139452P.
 PR 16-JUN-1999; 99US-0139453P.
 PR 17-JUN-1999; 99US-0139492P.
 PR 18-JUN-1999; 99US-0139454P.
 PR 18-JUN-1999; 99US-0139455P.
 PR 18-JUN-1999; 99US-0139456P.
 PR 18-JUN-1999; 99US-0139457P.
 PR 18-JUN-1999; 99US-0139458P.
 PR 18-JUN-1999; 99US-0139459P.
 PR 18-JUN-1999; 99US-0139460P.
 PR 18-JUN-1999; 99US-0139461P.
 PR 18-JUN-1999; 99US-0139462P.
 PR 18-JUN-1999; 99US-0139463P.
 PR 18-JUN-1999; 99US-0139750P.
 PR 18-JUN-1999; 99US-0139763P.
 PR 21-JUN-1999; 99US-0139817P.
 PR 22-JUN-1999; 99US-0139899P.
 PR 23-JUN-1999; 99US-0140353P.
 PR 23-JUN-1999; 99US-0140354P.
 PR 24-JUN-1999; 99US-0140695P.


```
OY 288 TRILONSLGNPKRIITCTTPV--SPDETITALQFASAKYKMTQYVNEVSTDEALLK 345
Db ||:||||: ||:||||| ||:||||| ||:||||| ||:||||| ||:||||| ||:|||||
359 TRLOSSLGHGHSVLSCTITPASSSSSEETHNTLKFSRAKSIYASRNQIIDEKSLIK 418
OY 346 RYRKEIMDLKKQLEEV-----SLETRAQAMEKQ--LAQLLEEKDLLOKQVNEK 392
Db ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:||||
419 KYQREISTLKLQDLQRGMVGVSHEELMSLKQLEEGOVQMSRLEEEERAKAALMSR 478
OY 393 IENITRLMLVTSS 404
Db ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:||||
479 IQRLTKLILVST 490

RESULT 14
AAG31111
ID AAG31111 standard; protein; 1069 AA.
XX
AC AAG31111;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37308.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134370P.
PR 19-MAY-1999; 99US-0134768P.
PR 20-MAY-1999; 99US-0134941P.
PR 21-MAY-1999; 99US-0135124P.
PR 24-MAY-1999; 99US-0135353P.
PR 25-MAY-1999; 99US-0135629P.
PR 27-MAY-1999; 99US-0136021P.
PR 28-MAY-1999; 99US-0136392P.
PR 01-JUN-1999; 99US-0136782P.
PR 03-JUN-1999; 99US-0137222P.
PR 04-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 21-JUN-1999; 99US-0139763P.
PR 22-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-014154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 21-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 26-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145911P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 04-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
```

Query Match	33.8%;	Score 694;	DB 3;	Length 1069;
Best Local Similarity	36.2%;	Pred. No. 6,7e-52;		
Matches 178; Conservative	87;	Mismatches 131;	Indels 96;	Gaps 14;

Fri Aug 6 10:49:09 2004

```
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 33.8%; Score 694; DB 3; Length 1121;
Best Local Similarity 36.2%; Pred. No. 7.2e-52;
Matches 178; Conservative 87; Mismatches 131; Indels 96; Gaps 14;

QY 1 AEGAVAVCVVRPLN-----SREES-----LGETA---OVYWKTDN 34
Db 94 SERDSISVTRFRPLRYARSDLAMDOQYCRDYSFHVDAIGVNSLLGEYQRGDEVANYPDG 153
QY 35 NVI--YQVDSKSFNDRVPHGNETTKNVYEEIAAPIIDSAIQYNGTTFPAYGQTASGKT 92
Db 154 DTLVRHEYNPLTAYAFDKVFPQATTIDVDVAARPVVKAAMEGVNGTVFAYGVTSSGKT 213
QY 93 YTM-----MGSEDLGVIPRAHDI 113
Db 214 HTMHVRLVKKKPYTPPSLVFLQNTCAMNDLFFNLVPLWYLLGDQSPGIIPLAIKOVF 273
QY 114 QTKKF-----PDREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVYVADLT 167
Db 274 SIIQDVSLNGTGTGREFLRVSYLEIYNEVINDLLDPTG--QNLVRRED--SQTYVEGK 330
QY 168 EYVYVTSSEMALKWTTKGKSRHYGETKMNQKSRSHITFRMILESRKGPSPNCEGSKV 227
Db 331 EEWLSPCHALSFAAGEEHRHVGSNNFNLLSSRSHITFTLMVRESSATGDEYD--GVIF 387
QY 228 SHLNLDLAGSRAAQTGAAGVRLKEGCNINRSFLIIGOVTKLSDQVGVGFNYRDSKL 287
Db 388 SOLNLDLAGSE--SSKETTGLRKEGKYINKSLTLTGTVIGKLGK--ATHIPYRDSKL 445
QY 288 TRILQNSLGNPKTRIICTITPV--SPDETLTALQFASAKYMKNTPYNEVSTDEALLK 345
Db 446 TRILQSSLSGHGVSLICTITTPASSSEETHTLKFPASRAKSIETIYASRNIIDKSLIK 505
QY 346 RYRKEINDLKKOLEV-----SLETRAQAMEKDQ--LAQLLEEKDLLQKVQNEK 392
Db 506 KYQREISTLKLELQDLRGMLVGVSHLEMSLKQOLEGQVKMQSRLEEEEEEAKAALMSR 565
QY 393 IENLRMLVTSS 404
Db 566 IQKLTKLILVST 577
```

Search completed: July 29, 2004, 09:35:13
Job time : 25.3176 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:30:02 ; Search time 7.15731 Seconds
(without alignments)
2914.068 Million cell updates/sec

Title: US-10-045-631b-88_COPY_2_405

Perfect score: 2053

Sequence: 1 AEEGAVACVVRPLNSREE.....LQKVQNEKIENLRMLVTSS 404

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pap:*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pap:*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pap:*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pap:*
5: /cgn2_6/ptodata/2/iaa/PCITUS COMB.pap:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2053	100.0	2662	4	US-09-595-684B-31
2	1468.5	71.5	2954	4	US-09-150-867-1
3	694.5	33.8	473	4	US-09-592-054-6
4	694.5	33.8	522	4	US-09-592-054-4
5	684.5	33.3	1231	4	US-09-595-684B-23
6	682.5	33.2	1234	4	US-09-592-054-8
7	678.5	33.0	1232	4	US-09-592-054-2
8	662.5	32.3	1066	3	US-09-541-782-8
9	662.5	32.3	1066	4	US-09-723-820-8
10	662.5	32.3	1066	4	US-10-270-085-8
11	652	31.8	411	2	US-08-713-815A-4
12	652	31.8	441	2	US-08-713-815A-3
13	652	31.8	975	4	US-09-914-259-19
14	652	31.8	1388	4	US-09-572-191-2
15	652	31.8	1388	4	US-09-723-262-2
16	652	31.8	1388	4	US-09-723-219-2
17	645	31.4	1031	4	US-09-914-259-24
18	637	31.0	967	4	US-09-914-259-21
19	634.5	30.9	1032	4	US-09-914-259-26
20	632	30.8	409	4	US-09-572-191-6
21	632	30.8	409	4	US-09-572-191-2
22	632	30.8	409	4	US-09-723-262-6
23	631.5	30.8	957	4	US-09-723-219-6
24	630.5	30.7	956	4	US-09-914-259-17
25	630.5	30.7	956	4	US-09-914-259-17
26	630.5	30.7	963	4	US-09-914-259-22
27	629.5	30.7	963	4	US-09-914-259-20

28	622	30.3	815	4	US-09-914-259-18	Sequence 18, Appl
29	609.5	29.7	928	4	US-09-914-259-23	Sequence 23, Appl
30	606.5	29.5	1056	4	US-09-595-684B-29	Sequence 29, Appl
31	606.5	29.5	1057	3	US-09-541-782-10	Sequence 10, Appl
32	606.5	29.5	1057	4	US-09-723-820-10	Sequence 10, Appl
33	606.5	29.5	1057	4	US-10-270-085-10	Sequence 2, Appl
34	605.5	29.5	1057	4	US-09-428-156B-2	Sequence 6, Appl
35	603.5	29.4	513	4	US-09-724-519-6	Sequence 6, Appl
36	603.5	29.4	513	4	US-09-592-037-6	Sequence 8, Appl
37	603.5	29.4	513	4	US-09-428-156B-6	Sequence 6, Appl
38	603.5	29.4	575	4	US-09-724-519-8	Sequence 8, Appl
39	603.5	29.4	575	4	US-09-592-037-8	Sequence 4, Appl
40	603.5	29.4	575	4	US-09-428-156B-8	Sequence 4, Appl
41	603	29.4	375	4	US-09-572-191-4	Sequence 25, Appl
42	603	29.4	375	4	US-09-723-262-4	Sequence 2, Appl
43	603	29.4	375	4	US-09-723-219-4	Sequence 2, Appl
44	603	29.4	935	4	US-09-914-259-25	Sequence 2, Appl
45	603	29.4	1279	4	US-09-724-517-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766

GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe

; APPLICANT: Ohashi, Cara

; APPLICANT: Sakowicz, Roman

; APPLICANT: Vaisberg, Eugeni

; APPLICANT: Wood, Kenneth

; APPLICANT: Yu, Ming

; TITLE OF INVENTION: Human kinesins and methods of producing
; FILE REFERENCE: cytop036

; CURRENT APPLICATION NUMBER: US/09/595,684B

; CURRENT FILING DATE: 2002-06-24

; PRIOR APPLICATION NUMBER: 09/295,612

; PRIOR FILING DATE: 2000-04-20

; NUMBER OF SEQ ID NOS: 105

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 31

; LENGTH: 2662

; TYPE: PRT

; ORGANISM: Human

US-09-595-684B-31

Query Match

Best Local Similarity 100.0%; Score 2053; DB 4; Length 2662;

Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	AEEGAVACVVRPLNSREE	SLGETAQVYWKTNVYQVDSKSFNDFRVFHNETHKN	60
Db	2	AEEGAVACVVRPLNSREE	SLGETAQVYWKTNVYQVDSKSFNDFRVFHNETHKN	61
Qy	61	VYEIAAPIIDSAIQYNGTIFAYGQTASGKTYTMGSEDLGVIPRAIHDIQKIKKFP	120	
Db	62	VYEIAAPIIDSAIQYNGTIFAYGQTASGKTYTMGSEDLGVIPRAIHDIQKIKKFP	121	
Qy	121	DREFLRVSYMEIYNETITDLGCTQMKPLIREDVNRNVYVADLITEEYVYVSEMAKWK	180	
Db	122	DREFLRVSYMEIYNETITDLGCTQMKPLIREDVNRNVYVADLITEEYVYVSEMAKWK	181	
Qy	181	ITKGEKSRHYGETKMNORSRSHITFRMTLESREKEPCNCEGSKVSHNLVLDLAGSER	240	
Db	182	ITKGEKSRHYGETKMNORSRSHITFRMTLESREKEPCNCEGSKVSHNLVLDLAGSER	241	
Qy	241	AAQTGAAGVRLKEGNCINRSLFILQVYKLSGQVGGFINTYRDSKLTILQNSLGNPK	300	
Db	242	AAQTGAAGVRLKEGNCINRSLFILQVYKLSGQVGGFINTYRDSKLTILQNSLGNPK	301	

Fri Aug 6 10:49:09 2004

QY 301 TRIICTTPVSPFDETLTALQFASAKYMKNTPYVNEVSTDEALLKRYKEIMDLKKQLEE 360
 Db 302 TRIICTTPVSPFDETLTALQFASAKYMKNTPYVNEVSTDEALLKRYKEIMDLKKQLEE 361
 QY 361 VSLERAQAMEKDQALAEKDLQKQVNEKIENLTRLMLVTSS 404
 Db 362 VSLERAQAMEKDQALAEKDLQKQVNEKIENLTRLMLVTSS 405

RESULT 2

US-09-150-867-1
 ; Sequence 1, Application US/09150867
 ; Patent No. 6645748
 ; GENERAL INFORMATION:
 ; APPLICANT: Wood, Kenneth W.
 ; APPLICANT: Sakowicz, Roman
 ; APPLICANT: Goldstein, Lawrence S.B.
 ; APPLICANT: Cleveland, Don W.
 ; APPLICANT: The Regents of the University of California
 ; TITLE OF INVENTION: Plus End-Directed Microtubule Motor Required for
 ; FILE REFERENCE: 18557C-000110US
 ; CURRENT APPLICATION NUMBER: US/09/150,867
 ; CURRENT FILING DATE: 1998-09-10
 ; EARLIER APPLICATION NUMBER: US 60/058,645
 ; EARLIER FILING DATE: 1997-09-11
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 2954
 ; TYPE: PRT
 ; ORGANISM: Xenopus sp.
 ; FEATURE:
 ; OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)
 ; OTHER INFORMATION: member of the kinesin superfamily of microtubule
 ; OTHER INFORMATION: motor proteins
 ; FEATURE:
 ; NAME/KEY: DOMAIN
 ; LOCATION: (1)..(472)
 ; OTHER INFORMATION: kinesin like motor domain
 ; FEATURE:
 ; NAME/KEY: DOMAIN
 ; LOCATION: (473)..(2752)
 ; OTHER INFORMATION: rod domain
 ; FEATURE:
 ; NAME/KEY: DOMAIN
 ; LOCATION: (2753)..(2954)
 ; OTHER INFORMATION: tail domain
 ; US-09-150-867-1

Query Match 71.5%; Score 1468.5; DB 4; Length 2954;
 Best Local Similarity 70.9%; Pred. No. 1.1e-127; Indels 7; Gaps 3;
 Matches 290; Conservative 52; Mismatches 60;
 QY 1 ABEGAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKGFNDRVPHGHETHKN 60
 Db 2 SEGDAVVCVVRPLIQREQ--GDQANLQKAGNNTISQVDTKGFNDRVFNESHSTQ 59
 QY 61 VYEEIAAPLIDSAIQYNGTTFAYGATAGKTYTMGSDHLGVIPRAITHDIFQKIKFP 120
 Db 60 IYCEIAVPIIRGALQYNGTTFAYGATAGKTYTMGSDHLGVIPRAITHDIFQKIKFP 119
 QY 121 DREFLLRVSMYMEIYNETITDLCGTQKMKPLIREDVNRNVYVADLTVEEVYVTSMAWK 180
 Db 120 NREFLLRVSMYMEIYNETITDLCDDRRKKPLEREDFNRVYVADLTVEELVMVPEHVIQ 179
 QY 181 ITKGEKSRHYGETKMNQSRSHITIFRMILESRKEGEPSS---NCEGSKVSHNLVLDLAG 237
 Db 180 IKKEGNRHYGETKMNQSRSHITIFRMIVESDRDPTNSNCDGAVVMVSHNLVLDLAG 239
 QY 238 SERAAQTGAAGVRLKEGNCNINRSLFILGVIKKLSGQVGFNRYRDSKLTILQNSLGG 297
 Db 240 SERASQTGABGVRLKEGNCNINRSLFILGVIKKLSGQAGGFINRYRDSKLTILQNSLGG 299

QY 298 NPKTRIICTTPVSPFDETLTALQFASAKYMKNTPYVNEVSTDEALLKRYKEIMDLKKQ 357
 Db 300 NAKTVIICTTPVSPFDETLTALQFASAKYMKNTPYVNEVSTDEALLKRYKEIMDLKKQ 359
 QY 358 LE--EVSLETRAQAMEKDQALAEKDLQKQVNEKIENLTRLMLVTSS 404
 Db 360 LENLESSETRAQAMEKKEHTQLLAIEIKQLHKEKREDRIWHLTNIVVASS 408
 RESULT 3
 US-09-592-054-6
 ; Sequence 6, Application US/09592054
 ; Patent No. 6440684
 ; GENERAL INFORMATION:
 ; APPLICANT: Beraud, Christophe
 ; APPLICANT: Finer, Jeffrey
 ; APPLICANT: Sakowicz, Roman
 ; APPLICANT: Wood, Kenneth
 ; TITLE OF INVENTION: No. 6440684el motor proteins and methods for
 ; TITLE OF INVENTION: their use
 ; FILE REFERENCE: 1016
 ; CURRENT APPLICATION NUMBER: US/09/592,054
 ; CURRENT FILING DATE: 2000-07-20
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 6
 ; LENGTH: 473
 ; TYPE: PRT
 ; ORGANISM: Human
 ; US-09-592-054-6

Query Match 33.8%; Score 694.5; DB 4; Length 473;
 Best Local Similarity 39.6%; Pred. No. 1.7e-56;
 Matches 169; Conservative 66; Mismatches 145; Indels 45; Gaps 8;
 QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKGFNDRVPHGHETHKN 64
 Db 8 VRVALRCPLVPKEISEGCMCLSFVPGEPQVVGTD--KSFTYDFVDFPSTQEVEVT 65
 QY 65 IAAPIIDSAIQYNGTTFAYGATAGKTYTMG-----SEDHLGVIPRAITHDIFQKIK 117
 Db 66 AVAPLIKGVFKGNATVLAGYQGTGSGKTYSMGGAYTAQENEPVGVIPVLIQLFKEID 125
 QY 118 KPPDREPLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVYVADLTVEEVYVTSMA 177
 Db 126 KKSDFEFTLVKVSYLEIYNEEILDLLCPSRKAQINREDPKGKIIVGLTEKTLVLDL 185
 QY 178 LKWITKGEKSRHYGETKMNQSRSHITIFRMILESRKEGEPSSNCEGSKVSHNLVLDLAG 237
 Db 186 VSCLEQGNNSRTVASTANSSQSRSHAFTTISLEQKESD---KNSSFRSKLHLVLDLAG 241
 QY 238 SERAAQTGAAGVRLKEGNCNINRSLFILGVIKKLSGQVGFNRYRDSKLTILQNSLGG 297
 Db 242 SERQKTKAEGDRLKEGINNRGLICLGNVISALGDDKGGFVPRYDSKLTILQNSLGG 301
 QY 298 NPKTRIICTITPVP--SFDETLTALQFASAKYMKNTPYVNEVSTDEALLKRYKEIMDLK 355
 Db 302 NSHTLMIACVSPADSNLEETLNTLAYADRAKIKKPKIVN-----IDPQTAENHLK 353
 QY 356 KQLEEVSL-----ETRAQAMEKDQALAEKDLQKQVNEKIENL 396
 Db 354 QQVQQLVLLQAHGGLPGSITVPFSPENLQSLMEKQ--SLVEENEKLSRGLSEAGQT 411
 QY 397 TRML 400
 Db 412 AQML 415

RESULT 4

US-09-592-054-4
 ; Sequence 4, Application US/09592054
 ; Patent No. 6440684

```

; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Finer, Jeffrey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6440684el motor proteins and methods for
; FILE REFERENCE: 1016
; CURRENT APPLICATION NUMBER: US/09/592,054
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Human
US-09-592-054-4

Query Match      33.8%; Score 694.5; DB 4; Length 522;
Best Local Similarity 39.6%; Pred. No. 2e-56;
Matches 168; Conservative 66; Mismatches 145; Indels 45; Gaps 8;

QY      6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDFVFGHNETTKNYEE 64
Db      30 VRVALRCRLVPKEISEGCMCLSFVPGEPQVVVGTD--KSFTYDFVDFPSTEQEEVNT 87
QY      65 IAAPIDSAIOGYNGTIFAYGQTASGKTYTMMG-----SEDLGVIPRAIHDFQKIK 117
Db      88 AVAPLIKGVFKYNATVLAYGQTGSKTYSMGAYTAEOENPTVGVIPRVQLLFKEID 147
QY      118 KPDPREFLLRVSYMEIYNETITDLCTGQKMKPLIREDVNRNVVADLTERVVYVTSEMA 177
Db      148 KKSDFEFTLKVSYLEIYNEEILDLLCPREKAQINIREDPKEGIKIVGLTEKTVLVALDT 207
QY      178 LKWTIKGKSRHYGKTKMQRSSRSHITFRMILESRKEGPNCEGSKVSHNLVLDLAG 237
Db      208 VSCLEQGNNSRTVASTAMNSQSSRSHAFTISLEQRKSD-----KNSFRSKLHLVDLAG 263
QY      238 SERAAQTAAGVRLKEGCNINRSLFILGVIKLSDGVGVGFYINVRDSKLTRELQNSLGG 297
Db      264 SERQKTKAEGDRLKEGININRGLCLGNVISALGDDKGGFAPYRDSKLTRELQNSLGG 323
QY      298 NPKTRIICITIPV--SPDETITAIQFASTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLK 355
Db      324 NSHTLMITACVSPADSNLEETLTLRYADRARKINKPIVN-----IDPQTAEHLNHLK 375
QY      356 KQLEEVSL-----ETRAQAMEKQDLAQLLEKCOLLQKVKQNEKIENL 396
Db      376 QVQVQLQVLLQLQAHGGLTPGSITVEPSENLOSMEKNQ--SLVBEENEKLSRGLSEAAAGT 433
QY      397 TRML 400
Db      434 AQML 437

RESULT 5
US-09-595-684B-23
; Sequence 23, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; FILE REFERENCE: Cyt0036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20

; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Finer, Jeffrey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6440684el motor proteins and methods for
; FILE REFERENCE: 1016
; CURRENT APPLICATION NUMBER: US/09/592,054
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1234
; TYPE: PRT
; ORGANISM: Human
US-09-592-054-8

Query Match      33.2%; Score 682.5; DB 4; Length 1234;
Best Local Similarity 38.8%; Pred. No. 9.8e-55;
Matches 166; Conservative 73; Mismatches 142; Indels 49; Gaps 10;

QY      6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDFVFGHNETTKNYEE 64

```

238 SERAQTGAAGVRLKEGNCINRSLFILGQVVKLSLSDGQVGGFVNYRDSKLTILQNSLGG 297
 244 SERQKTKAEGDRLKEGINRGLLCLGNVISALGDDKGGFAPYRDSKLTLLQSLGG 303
 298 NPKTRIICTITPV--SFDETLTALQFASAKYMKNTPPYNEVSTDEALLKRYRKEIMDLK 355
 304 NSHTLMIAVSPADSNLEETLTLRYADRARKIKNKPVN-----IDPQTAELNHLK 355
 356 KQLEEVSL-----ETRAQAMEKDQALAEKDLQKQVNEKIENL 396
 356 QVQQLQVLLQAHGGLPGSITVPESENLSLMEKNQ--SLVEENEKLSRGLSEAGQT 413
 397 IEML 400
 414 AQML 417

RESULT 8
 US-09-541-782-8
 ; Sequence 8, Application US/09541782
 ; Patent No. 6284480
 ; GENERAL INFORMATION:
 ; APPLICANT: Nislow, Corey
 ; APPLICANT: Sakowicz, Roman
 ; APPLICANT: Beraud, Christophe
 ; TITLE OF INVENTION: Antifungal assay
 ; FILE REFERENCE: 1015
 ; CURRENT APPLICATION NUMBER: US/09/541,782
 ; CURRENT FILING DATE: 2000-04-03
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 8
 ; LENGTH: 1066
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 US-09-541-782-8

Query Match 32.3%; Score 662.5; DB 3; Length 1066;
 Best Local Similarity 37.4%; Pred. No. 5.8e-53;
 Matches 159; Conservative 78; Mismatches 141; Indels 47; Gaps 11;

QY 6 VAVCVVRPLNSREESL--GETAQVYWKTDNNVIYQVDG--SKSPNDRVPHGNETTKNV 61
 DB 20 IQVTVVRPLNSREERCIRSAEVDVVGPREVWTRHTLDSKLTKKFTFDRSFGPSKQCDV 79
 QY 62 YEEIAPIIDSAIOQYNGTIFAYGQTASGKTYTMGSE-----DHLGVIPRAIH 110
 DB 80 YSVVVSPLIEVLNGYNTVFAYGQTGTGKTHVWGNETAELKSWEDSDIGIIPRAIS 139
 QY 111 DIFQKIKFPDREFLLRVSVMEIYNETITDLCGTQRMKPLIREDVNR--VYVADLTEE 169
 DB 140 HLFDELMM--EYETMRISYLELYNEELCDLLSTDDTKIRIFDDSTKKGSVIIQGLEEI 198
 QY 170 VVYVSEMAKWTIKGEKSRHYGETMQRSSRSHITFRMILESREKGPSPNCEGSKVYSH 229
 DB 199 PVHSKDDVYKLEKGERKKTATILMAQSSRSHTVFSIVVHIRENGIEG--EDMLKIGK 256
 QY 230 LNLVDLAGSRAAQTG--AAGVRLKEGNCINRSLFILGQVVKLSLSDGQVGGFVNYRDSKLT 288
 DB 257 LNLVDLAGSENKAGNEKIRVRETVMNOSLLTLGRVITLVDR--APHVPYRESKLT 314
 QY 289 RILQNSLGNPKTRIICTITP--VSFDETLTALQFASAKYMKNTPPYNEVSTDEALLK 346
 DB 315 RILQNSLGRYKTSIIATISPHGKDIETLSTLEYAHRANKIQNKPEVNQKLTTKTVLKE 374
 QY 347 YRKEIMDLKQI-----EVSLETRAQAMEKDQALAEKDLQKQVNEKIENL 390
 DB 375 YTEEDIKRDLMAARDKNGIYLABETYGEITLKESQNR-----LNEKMLLLKALK 427
 QY 391 EKIN 395
 DB 428 DELQN 432

10 VRVALRCRPLVPKEISEGQCLSFVPGETQVVVGTD--KSFYDFVDPCTEQEVEFNK 67
 65 IAAPLIDSAIOQYNGTIFAYGQTASGKTYTMG-----SEDHLGVIPRAIHDFQKIK 117
 68 AVAPLIKGFKNATVLAAYGQSGKTYSMGAYTAEQENEPVGLIPRVILQLLKFKEID 127
 118 KFPDREFLLRVSVMEIYNETITDLCGTQKMKPLIREDVNRNVVADLTEEVVYTSEMA 177
 128 QKSDFEFTLKVSYLEIYNEEILDLCPREKAQINIREDPKGIKIVGLTEKTVLVALDT 187
 178 LKWITKGEKSRHYGETMQRSSRSHITFRMILESREKGPSPNCEGSKVSHNLNVDLAG 237
 188 VSLCQGNNSRIVASTAMNSQSSRSHAIPTI--SLEQKKSDKNSFR--SKLHLVDLAG 243
 238 SERAQTGAAGVRLKEGNCINRSLFILGQVVKLSLSDGQVGGFVNYRDSKLTILQNSLGG 297
 244 SERQKTKAEGDRLKEGINRGLLCLGNVISALGDDKGGFAPYRDSKLTLLQSLGG 303
 298 NPKTRIICTITPV--SFDETLTALQFASAKYMKNTPPYNEVSTDEALLKRYRKEIMDLK 355
 304 NSHTLMIAVSPADSNLEETLTLRYADRARKIKNKPVN-----IDPHTAELNHLK 355
 356 KQLEEVSL-----ETRAQAMEKDQALAEKDLQKQVNEKIENL 396
 356 QVQQLQVLLQAHGGLPGSITVPESENLSLMEKNQ--SLVEENEKLSRGLSEAGQT 413
 393 IEMLTMLVT 402
 414 AQMLERIIILT 423

RESULT 7
 US-09-592-054-2
 ; Sequence 2, Application US/09592054
 ; Patent No. 6440684
 ; GENERAL INFORMATION:
 ; APPLICANT: Beraud, Christophe
 ; APPLICANT: Finer, Jeffrey
 ; APPLICANT: Sakowicz, Roman
 ; APPLICANT: Wood, Kenneth
 ; TITLE OF INVENTION: No. 6440684el motor proteins and methods for
 ; TITLE OF INVENTION: their use
 ; FILE REFERENCE: 1016
 ; CURRENT APPLICATION NUMBER: US/09/592,054
 ; CURRENT FILING DATE: 2000-07-20
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 1232
 ; TYPE: PRT
 ; ORGANISM: Human
 US-09-592-054-2

Query Match 33.0%; Score 678.5; DB 4; Length 1232;
 Best Local Similarity 39.4%; Pred. No. 2.3e-54;
 Matches 167; Conservative 66; Mismatches 146; Indels 45; Gaps 9;

QY 6 VAVCVVRPLNSREESL--GETAQVYWKTDNNVIYQVDGSKSPNDRVPHGNETTKNVYEE 64
 DB 10 VRVALRCRPLVPKEISEGQCLSFVPGEPQVVVGTD--KSFYDFVDPCTEQEVEFNK 67
 QY 65 IAAPLIDSAIOQYNGTIFAYGQTASGKTYTMG-----SEDHLGVIPRAIHDFQKIK 117
 DB 68 AVAPLIKGFKNATVLAAYGQSGKTYSMGAYTAEQENEPVGLIPRVILQLLKFKEID 127
 QY 118 KFPDREFLLRVSVMEIYNETITDLCGTQKMKPLIREDVNRNVVADLTEEVVYTSEMA 177
 DB 128 KKSDFEFTLKVSYLEIYNEEILDLCPREKAQINIREDPKGIKIVGLTEKTVLVALDT 187
 QY 178 LKWITKGEKSRHYGETMQRSSRSHITFRMILESREKGPSPNCEGSKVSHNLNVDLAG 237
 DB 188 VSLCQGNNSRIVASTAMNSQSSRSHAIPTI--SLEQKKSDKNSFR--SKLHLVDLAG 243

```

; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-270-085-8

Query Match          32.3%; Score 662.5; DB 4; Length 1066;
Best Local Similarity 37.4%; Pred. No. 5.8e-53;
Matches 159; Conservative 78; Mismatches 141; Indels 47; Gaps 11;

QY      6 VAVCVVRPLNSRRESL--GETAQVYWKTDNNVYIQVDG--SKSFNDRVPHGNETTKNV 61
DB      20 IQVYVRPLNSRRCIRSAEVDVGVREVVTTRHTLDSKLTKKFTFDRSGPESKQCDV 79

QY      62 YEEIAAPIIDSIAIQYNGTIFAYGOTAGKTYTMWGSE-----DHLGVIPRAIH 110
DB      80 YSVVVSPLIEVLNGYNCVTFAYGOTGTGKHTVWGNETAELKSSWEDSDGIHPRALS 139

QY      111 DIFQIKKPPDRFLLRVSYMEIYNETITDLCGTQMKPLIIREVDYNR--NVTVADLTEE 169
DB      140 HLFDELRRM--EVEYTRMISVLEYNEELCDLLSTDYTKIRIPDDSTKGSVIOGLEEI 198

QY      170 VVYTESMALKWITTKGSRHVGETKMQRSRSHITFRMILEGREKEPNCCEGSKVKVSH 229
DB      199 PVHSKDDVYKLLGKGRKRTATTFLMNAQSSRSHITFVSIVVHIRENGIEG--EDMLKIGK 256

QY      230 LNLVDLAGSBRAAQTG--AAGVRLKEGCNINRSLFILGOVTKKLSDGOVGGFYINVRDKLT 288
DB      257 LNLVDLAGSENVSKAGNEKGI RVETVNIQSLTLGRVITALVDLR--APHVPIRESKLT 314

QY      289 RILNSLGGNPKTRIICTTTP--VSFDETLTALQFASATKYMKNTPVYNEVSTDEALLKR 346
DB      315 RLLQESLGGRTKTSIIATISPGHKDTEETLSTLEYAHRAKNIQKPEVQNKLTKTVLKE 374

QY      347 YRKEIMDLKKQL-----EVSLETRAQAMEKDLQAQLLEEKDLLOKVQN 390
DB      375 YTEIDKLRDLMAARDKNGIYLAEEITYGEITIKLESQNR-----LNEKMLLLKALK 427

QY      391 EKIEI 395
DB      428 DELQN 432

RESULT 11
US-08-713-815A-4
; Sequence 4, Application US/08713815A
; Patent No. 5830659
; GENERAL INFORMATION:
; APPLICANT: Russell J. Stewart
; TITLE OF INVENTION: ACTIVE MICROTUBULE-BASED
; TITLE OF INVENTION: SEPARATIONS BY KINESINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thorpe, No. 5830659th & Western, L.L.P.
; STREET: 9035 South 700 East, Suite 200
; CITY: Sandy
; STATE: Utah
; COUNTRY: USA
; ZIP: 84070
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: AST Ascentia 900N
; OPERATING SYSTEM: DOS 6.22
; SOFTWARE: Word Perfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08713,815A
; FILING DATE: 13-SEP-1996
; CLASSIFICATION: 435

```

Fri Aug 6 10:49:09 2004

ATTORNEY/AGENT INFORMATION:
NAME: Alan J. Howarth
REGISTRATION NUMBER: 36,553
REFERENCE/DOCKET NUMBER: T3214/U-2202
TELECOMMUNICATION INFORMATION:
TELEPHONE: (801)566-6633
TELEFAX: (801)566-0750
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 441 amino acid residues
TYPE: amino acid
TOPOLOGY: linear
US-08-713-815A-3

Query Match 31.8%; Score 652; DB 2; Length 441;
Best Local Similarity 42.5%; Pred. No. 1.5e-52;
Matches 174; Conservative 58; Mismatches 149; Indels 28; Gaps 15;

QY 1 AEGGAVAVCVVRPLNSREESLGETAQVYKWTNNV---IYQVDGSKSFNDFRVFHGNET 57
Db 9 AEDSIKVCV-RFRPLNDSSEKAGSKFV--KFPNNVEENCISIAAG-KVYLFDRKVKFNAS 64
QY 58 TKNVYEIAPIIDSALQGYNGTIFAYGQTASGKTYM---MGSEDLHGVIPRAIHDFQ 114
Db 65 QEKVYNEAAKSIYDVLGAGYNGTIFAYGQTSSGKTHMEGVIGDSVKQGIIPRVNDIFN 124
QY 115 KIKKFP--DREFLLRVSYMEIYNETITDLCGTQMKPLIREDVNRNYYVADLVEEVVYT 173
Db 125 HIYAMENVLEPHIKVSYEYIMDKIRLL--DVSQVN--LSVHEDKRVYVKGATERFVSS 182
QY 174 SEMALKWITKGEKSRHYGETMQRNORSRSHITFRMILESREKGPSCGSKVYS-HLNL 232
Db 183 PEDVFEVIEEGKSNRHHIAVTNNHSSRSHSVFLNVKQ-----ENLENQKLSGKLYL 236
QY 233 VDLASERAAQTGAAGVRLKEGCNINRSLFTLGQVVKLSQGVGGFINYRDSKLTILQ 292
Db 237 VDLAGEKVKSTGAGTVDLDAKNINKLSALGNVISALADGN-KTHIPYRDSKLTILQ 295
QY 293 NSLGNPKTRICTITPVSF--ETLTALQFASTAKYMKNTPVYNEVSTDEALLKRYKE 350
Db 296 ESLGNARTTIVICCPASPFNESETKSTLDFGRRAKTVKNVVCVNEELTAEWKRYEKE 355
QY 351 ---IMDLKKQLEVSLE--TRAQMEKQOLAELLEKDLIQ-KVQNEKIE 394
Db 356 KEKNARLKGKVEKLEIARWRAGETVKAEEQINMEDLMEASTPNLEVE 404

RESULT 13
US-09-914-259-19
Sequence 19, Application US/09914259
Patent No. 6495336
GENERAL INFORMATION:
APPLICANT: Makowski, Lee
APPLICANT: Hyman, Paul
APPLICANT: Williams, Mark
TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
FILE REFERENCE: 8471-010-999
CURRENT APPLICATION NUMBER: US/09/914,259
CURRENT FILING DATE: 2000-11-21
NUMBER OF SEQ ID NOS: 180
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 19
LENGTH: 975
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-914-259-19

Query Match 31.8%; Score 652; DB 4; Length 975;
Best Local Similarity 42.5%; Pred. No. 4.9e-52;
Matches 174; Conservative 58; Mismatches 149; Indels 28; Gaps 15;

QY 1 AEGGAVAVCVVRPLNSREESLGETAQVYKWTNNV---IYQVDGSKSFNDFRVFHGNET 57
Db 9 AEDSIKVCV-RFRPLNDSSEKAGSKFV--KFPNNVEENCISIAAG-KVYLFDRKVKFNAS 64
QY 58 TKNVYEIAPIIDSALQGYNGTIFAYGQTASGKTYM---MGSEDLHGVIPRAIHDFQ 114
Db 65 QEKVYNEAAKSIYDVLGAGYNGTIFAYGQTSSGKTHMEGVIGDSVKQGIIPRVNDIFN 124
QY 115 KIKKFP--DREFLLRVSYMEIYNETITDLCGTQMKPLIREDVNRNYYVADLVEEVVYT 173
Db 125 HIYAMENVLEPHIKVSYEYIMDKIRLL--DVSQVN--LSVHEDKRVYVKGATERFVSS 182
QY 174 SEMALKWITKGEKSRHYGETMQRNORSRSHITFRMILESREKGPSCGSKVYS-HLNL 232
Db 183 PEDVFEVIEEGKSNRHHIAVTNNHSSRSHSVFLNVKQ-----ENLENQKLSGKLYL 236
QY 233 VDLASERAAQTGAAGVRLKEGCNINRSLFTLGQVVKLSQGVGGFINYRDSKLTILQ 292
Db 237 VDLAGEKVKSTGAGTVDLDAKNINKLSALGNVISALADGN-KTHIPYRDSKLTILQ 295
QY 293 NSLGNPKTRICTITPVSF--ETLTALQFASTAKYMKNTPVYNEVSTDEALLKRYKE 350
Db 296 ESLGNARTTIVICCPASPFNESETKSTLDFGRRAKTVKNVVCVNEELTAEWKRYEKE 355
QY 351 ---IMDLKKQLEVSLE--TRAQMEKQOLAELLEKDLIQ-KVQNEKIE 394
Db 356 KEKNARLKGKVEKLEIARWRAGETVKAEEQINMEDLMEASTPNLEVE 404

ATTORNEY/AGENT INFORMATION:
NAME: Alan J. Howarth
REGISTRATION NUMBER: 36,553
REFERENCE/DOCKET NUMBER: T3214/U-2202
TELECOMMUNICATION INFORMATION:
TELEPHONE: (801)566-6633
TELEFAX: (801)566-0750
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 411 amino acid residues
TYPE: amino acid
TOPOLOGY: linear
US-08-713-815A-4

Query Match 31.8%; Score 652; DB 2; Length 411;
Best Local Similarity 42.5%; Pred. No. 1.3e-52;
Matches 174; Conservative 58; Mismatches 149; Indels 28; Gaps 15;

QY 1 AEGGAVAVCVVRPLNSREESLGETAQVYKWTNNV---IYQVDGSKSFNDFRVFHGNET 57
Db 9 AEDSIKVCV-RFRPLNDSSEKAGSKFV--KFPNNVEENCISIAAG-KVYLFDRKVKFNAS 64
QY 58 TKNVYEIAPIIDSALQGYNGTIFAYGQTASGKTYM---MGSEDLHGVIPRAIHDFQ 114
Db 65 QEKVYNEAAKSIYDVLGAGYNGTIFAYGQTSSGKTHMEGVIGDSVKQGIIPRVNDIFN 124
QY 115 KIKKFP--DREFLLRVSYMEIYNETITDLCGTQMKPLIREDVNRNYYVADLVEEVVYT 173
Db 125 HIYAMENVLEPHIKVSYEYIMDKIRLL--DVSQVN--LSVHEDKRVYVKGATERFVSS 182
QY 174 SEMALKWITKGEKSRHYGETMQRNORSRSHITFRMILESREKGPSCGSKVYS-HLNL 232
Db 183 PEDVFEVIEEGKSNRHHIAVTNNHSSRSHSVFLNVKQ-----ENLENQKLSGKLYL 236
QY 233 VDLASERAAQTGAAGVRLKEGCNINRSLFTLGQVVKLSQGVGGFINYRDSKLTILQ 292
Db 237 VDLAGEKVKSTGAGTVDLDAKNINKLSALGNVISALADGN-KTHIPYRDSKLTILQ 295
QY 293 NSLGNPKTRICTITPVSF--ETLTALQFASTAKYMKNTPVYNEVSTDEALLKRYKE 350
Db 296 ESLGNARTTIVICCPASPFNESETKSTLDFGRRAKTVKNVVCVNEELTAEWKRYEKE 355
QY 351 ---IMDLKKQLEVSLE--TRAQMEKQOLAELLEKDLIQ-KVQNEKIE 394
Db 356 KEKNARLKGKVEKLEIARWRAGETVKAEEQINMEDLMEASTPNLEVE 404

RESULT 12
US-08-713-815A-3
Sequence 3, Application US/08713815A
Patent No. 5830659
GENERAL INFORMATION:
APPLICANT: Russell J. Stewart
APPLICANT: Russell J. Stewart
TITLE OF INVENTION: ACTIVE MICROBULE-BASED
TITLE OF INVENTION: SEPARATIONS BY KINESINS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thorpe, No. 5830659th & Western, L.L.P.
STREET: 9035 South 700 East, Suite 200
CITY: Sandy
STATE: Utah
COUNTRY: USA
ZIP: 84070
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
COMPUTER: AST Ascentia 900N
OPERATING SYSTEM: DOS 6.22
SOFTWARE: Word Perfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/713,815A
FILING DATE: 13-SEP-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

us-10-045-631b-88_copy_2_405.ra1

Fri Aug 6 10:49:09 2004

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:39:47 ; Search time 20.1784 Seconds
(without alignments)
6280.361 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_405

Perfect score: 2053

Sequence: 1 ABEAVAVCVRVPLNSREE.....LQVONKEIENLRLMTVSS 404

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap.*
2: /cgn2_6/ptodata/2/pubpaa/PTC_NEW_PUB.pap.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pap.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pap.*
6: /cgn2_6/ptodata/2/pubpaa/PTCUS_PUBCOMB.pap.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pap.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pap.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pap.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pap.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pap.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pap.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pap.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pap.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	927	45.2	1382	16	US-10-437-963-176714
2	872.5	42.5	694	12	US-10-425-114-59725
3	744.5	36.3	1006	16	US-10-437-963-147933
4	718.5	35.0	420	16	US-10-437-963-114373
5	708.5	34.5	956	16	US-10-437-963-182113
6	705.5	34.4	955	12	US-10-425-114-62672
7	694.5	33.8	1232	14	US-10-116-712-670
8	694.5	33.8	1232	16	US-10-408-765A-2153
9	694.5	33.8	1235	15	US-10-334-143-8
10	691.5	33.8	1237	15	US-10-334-143-33
11	691.5	33.7	965	16	US-10-437-963-169150
12	691.5	33.7	1029	16	US-10-311-642-2
13	684.5	33.3	1232	14	US-10-116-712-664
14	684.5	33.3	1232	14	US-10-116-712-669
15	682.5	33.2	672	16	US-10-408-765A-1664

16	671.5	32.7	966	12	US-10-425-114-62697
17	663.5	32.3	406	15	US-10-369-493-3789
18	660.5	32.2	517	15	US-10-369-493-12881
19	652	31.8	975	14	US-10-080-608A-19
20	652	31.8	975	15	US-10-370-685-108
21	652	31.8	1388	12	US-10-332-089-2
22	652	31.8	1388	14	US-10-146-473-82
23	652	31.8	1388	15	US-10-173-993-32
24	652	31.8	1388	16	US-10-188-832-164
25	651.5	31.7	757	16	US-10-220-120-366
26	651.5	31.7	757	16	US-10-363-829-316
27	645	31.4	1031	14	US-10-080-608A-24
28	645	31.4	1031	15	US-10-370-685-113
29	644.5	31.4	1022	16	US-10-437-963-133108
30	644.5	31.4	1362	14	US-10-296-838-2
31	643.5	31.3	1826	16	US-10-408-765A-1477
32	643	31.3	821	16	US-10-437-963-114370
33	637	31.0	967	14	US-10-080-608A-21
34	637	31.0	967	15	US-10-370-685-110
35	634.5	30.9	1011	12	US-10-336-472-24
36	634.5	30.9	1011	12	US-10-236-417-58
37	634.5	30.9	1032	14	US-10-080-608A-26
38	634.5	30.9	1032	15	US-10-370-685-115
39	632	30.8	409	12	US-10-332-089-6
40	631.5	30.8	957	14	US-10-080-608A-16
41	631.5	30.8	957	15	US-10-370-685-105
42	631.5	30.8	1027	14	US-10-080-608A-27
43	631.5	30.8	1027	15	US-10-370-685-116
44	630.5	30.7	956	14	US-10-080-608A-17
45	630.5	30.7	956	15	US-10-370-685-106

ALIGNMENTS

RESULT 1

US-10-437-963-176714

; Sequence 176714, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 176714

; LENGTH: 1382

; TYPE: ERT

; ORGANISM: Oryza sativa

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT4530_74437C.1.pap

US-10-437-963-176714

Query Match 45.2%; Score 927; DB 16; Length 1382;
Best Local Similarity 51.6%; Pred. No. 4.9e-77;
Matches 213; Conservative 62; Mismatches 108; Indels 30; Gaps 11;

Qy 6 VAVCVRVPLNSREESIGETAQYVWKTDDNVI-YQVDSKSFNDRVPHGNETTXYVEE 64

Db 4 IHVAVRARPITA--EDAGSP---WRVSGNAIALSTQPSIRFEFDIFGECRTADYVGA 58

Qy 65 IAAPIDSAIQYNGTIFAYGQTASGKTYTMGSEDLHGVIPRAIHDIQKIKFPDEF 124

Db 59 RTKHIVDSVARGNGTVFAVQVNSGKTYMRSNGEPGIIPLAVHDLFRTIEHLDREF 118

Db 327 ALGNKTSIICTAEPETHIEETRTLOFASRAKVCNCAQVNEILTDAILKQKLEI 386
 QY 352 MDLKKQ-----LEVSLETR-----AQAMEKDQALAEKDLQKQVNEKIENTERM 399
 Db 387 BELRKLQSHSEGLEQVVLKLRNDMHKSELERDLAMELEDERKLRMTLEQHLTEQOKL 446
 QY 400 LVTSS 404
 Db 447 EAISS 451

RESULT 3

US-10-437-963-147933
 ; Sequence 147933, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 147933
 ; LENGTH: 1006
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1)...(1006)
 ; OTHER INFORMATION: unsure at all Xaa locations
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_48414C.1.pep
 ; US-10-437-963-147933

Query Match 36.3%; Score 744.5; DB 16; Length 1006;
 Best Local Similarity 41.7%; Pred. No. 3.7e-60;
 Matches 178; Conservative 84; Mismatches 134; Indels 31; Gaps 12;

QY 1 AEGAVAVCVVRPLNREESLGETAQVYWKTDNNVYQVDSK-SFNFDVFGHNETK 59
 Db 64 AAKESVAVAVRPLSPREVRRGE--KIAYADGETVARSEQSNLAYAYDRVFGPTTTR 121
 QY 60 NVYEIAAPIIDSALQINGTIFAYGQTASGKTYTMMGSEDLGVIPRAIHDFOKIKKF 119
 Db 122 HYDVAQVYVNGAMKINGTIFAYGVTSKTHMGDQSPGVIPLAVKDFINIOET 181
 QY 120 PDREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNRYVADLTREVVYSEMAIK 179
 Db 182 PNREFLLRVSYLEIYNEVNDLL--NPAGNLRIRIEDLQGTI-VEGIKEEAVLSPVHALS 238
 QY 180 WITKGE---KSRHYGETKONQRRSSHTIFRMILESREKPEPSNCEGSKVYSHLNLVLA 236
 Db 239 LIAAGEVSELRHVSTNENLSSRSHTIFTLTIESPRGOSNEAE-AVTLSQLNLIDIA 297
 QY 237 GSERAQTAAGVRLKEGNCINRSLFILGOVTKLSDGQGVGFNVRDSKLTIRILQNSIG 296
 Db 298 GSE-SSRVETAGVHOKEGSYINKSLTLTKGVLSKLT-DKATHI9FRDSKLTLLKSSLS 355
 QY 297 GNPKTRITICTIPVS--FDETLTALQFATAKYMKNTPVYNEVSTDEALLKRYKEIMDL 354
 Db 356 GQGRVSLICTVTPASSNSEETHNTLKFAHRAKHIEIQTAKQNKIMDARSLLIKKYQNEIROL 415
 QY 355 KQOLEVSELETAQAMEKDQALQ---LLEEK-DLLQKVQ-----EKENTLT 397
 Db 416 KEELEQLRRSIRTGPTIEDTMQKHLLTETKEDCNVKLSRLEQGEAKAALLERIEHLT 475

QY 125 LLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNRYVADLTREVVYSEMAIKWITKG 184
 Db 119 LLRMSYMEIYNEIENDLL--VPEHRKLQIHESIERGIYVAGLRREIIVTCPEQVLEPMSFG 176
 QY 185 EKSRHYGETKONQRRSSHTIFRMILESREKPEPS---NCEGSKVYSHLNLVLDLAGSER 240
 Db 177 ESHRHIGETNMVYSSRSHTIFRMVIESREKVDSEBAGESCD-AVRVSVNLVLDLAGSER 235
 QY 241 AAGTGAAGVRLKEGNCINRSLFILGOVTKLSDG--QVGFNRYRDSKLTIRILQNSLGN 298
 Db 236 AAKTGAAGVRLKEGSHINKSLMTGLTVIKLSREGQGGHVPYRDSKLTIRILQNALGN 295
 QY 299 PKTRIICTIT--PVSPDETLTALQFATAKYMKNTPVYNEVSTDEALLKRYKEIMDLKK 356
 Db 296 ANTAIICNTITLAQVHADETKSSLOFASRALRVNACVNEILLDAALLKQKREIBELRA 355
 QY 357 QL-----BEVSELETAQAMEKDQALAEKDLQKQVNEKIENTERMVITS 403
 Db 356 KLRSELEKEKESISLE-----LLEEKKAKEQDKELIE--QAKKIENTLSSVLNS 401

RESULT 2

US-10-425-114-59725
 ; Sequence 59725, Application US/10425114
 ; Publication No. US20040034888A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Jingdong
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Screen, Steven E
 ; APPLICANT: Tabaska, Jack E
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53313)B
 ; CURRENT APPLICATION NUMBER: US/10/425,114
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 73128
 ; SEQ ID NO 59725
 ; LENGTH: 694
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: LIB3912-011-C4_FLI.pep
 ; US-10-425-114-59725

Query Match 42.5%; Score 872.5; DB 12; Length 694;
 Best Local Similarity 46.4%; Pred. No. 2.2e-72;
 Matches 197; Conservative 71; Mismatches 128; Indels 29; Gaps 9;

QY 6 VAVCVVRP-----LNSREESLGETAQVYWKTDN---NVYQVDG---SKSFNFDVFGH 54
 Db 30 ISVAVRPPNPLVAANTSPASSGGGDRWRIDDTRVSLHRAAGPITGASFPHVFDG 89
 QY 55 NETTKNVEEIAAPITDSALQYNGTIFAYGQTASGKTYTMMGSEDLGVIPRAIHDFIQ 114
 Db 90 AANNERIYGTVREILGAVGGFNGTAFAYGTSKGTFTMNGSDADPGIIPRAVRVFD 149
 QY 115 KIKFPDREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNRYVADLTREVVTS 174
 Db 150 TVRQADREFLLRVSYMEIYNEIENDLL--TLEGQKLKHESLDRGVYVSGLRREIVNSA 207
 QY 175 EMALKWITKESRHYGETKONQRRSSHTIFRMILESREKPEPSNCEGSKVYSHLNLVD 234
 Db 208 EQVFELLQGEANRHFGETNMNRSRSHITFRWLESSEKQDTQCGD-AIRVSVNLVD 266
 QY 235 LAGSRAAQTGAAGVRLKEGNCINRSLFILGOVTKLSD-QVGGFYNRYRDSKLTIRILQN 293
 Db 267 LAGSERIITGAGVRLNEGKYINKSLMILGNVINKLSENGKQKHIPYRDSKLTIRLOP 326
 QY 294 SLGGNPKTRITICTITP--VSFDETLTALQFATAKYMKNTPVYNEVSTDEALLKRYKEI 351

QY 398 RMLVTS 404
Db 476 ELILVA 482

RESULT 4

US-10-437-963-114373
; Sequence 114373, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437, 963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 114373
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; LOCATION: (1)..(956)
; OTHER INFORMATION: unsure at all Xaa locations
; OTHER INFORMATION: Clone ID: PAT_MRT4530_18070C.1.pap
US-10-437-963-114373

Query Match 35.0%; Score 718.5; DB 16; Length 420;
Best Local Similarity 41.9%; Pred. No. 2.6e-58;
Matches 169; Conservative 72; Mismatches 129; Indels 33; Gaps 9;

QY 6 VAVCVVRPLNSREESLGETAQVYKTDNNVIYQVDS-----KSFNFDVPHGNET 57
Db 17 IVSVRLRPVNAEARGDSD--WECAGPTTLFRGAVPERAFPASYSYDRVFSHECG 74
QY 58 TKNVVEIAAIIISATOGYNGTIFAYQTSAGTKTYTMGSEDHGVIPRAIHDFQIK 117
Db 75 TRQYDSGARQVAMSVLSGINASIFAYQTSAGTKTYTMGITEY-----SMSDIYDIE 128
QY 118 KFPDFEFLRYSMEIYNETITDLCGQTKMKPLIREDVNRNVVADLVEEVYVTSMA 177
Db 129 KHPREFILFSAEYIYNEAVRDLL--SSDATPLRLDDPEKGVVPEKLTETLRDKGHL 186
QY 178 LKWTKEKSRHYGETKMNQSRSSHTIFRMILESREKGEPSNCEGSKVYSHNLNLDV 237
Db 187 LELAVCEAQRIGETAMNEASSRSHQILRMVTESSAKQFLGKNSLTIAVNFVLDAG 246
QY 238 SERAAQTGAAGVRLKEGCNINRSFILQGVTKLSDGQGVGFYINRDSKLTIRLQNSLGG 297
Db 247 SERASQTASAGRLKEGSHINRSLLTTLGKVIQRLQSKGR--NGHIPYRDSKLTIRLQSSLGG 305
QY 298 NPKTRIICTIPV--SPDETALQFASAKYMKNTFYVNEVSTDEALLKRYKEIMDLK 355
Db 306 NARTALICTWSPACHIEQSRNTLLFANCAKDVVNAQVNVVMSDQALVKHLOREIARLE 365
QY 356 KQLE---EVSLTEAQAAM-EKD-----QLAQLLEEKDLLQ 386
Db 366 NELKFPASASCTSHAEILREKDELKLNLEQLMEQKDTVQ 408

RESULT 5

US-10-437-963-182113
; Sequence 182113, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437, 963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 182113
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(956)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_79331C.1.pap
US-10-437-963-182113

Query Match 34.5%; Score 708.5; DB 16; Length 956;
Best Local Similarity 42.3%; Pred. No. 8e-57;
Matches 176; Conservative 73; Mismatches 138; Indels 29; Gaps 11;

QY 2 BEGAVAVCVVRPLNSREESLGETAQVYKTDNN--VIYQVD-----GSKGFNFDVPHG 54
Db 31 KEKIFVTVRVRPLSKKELAKD--QVAMECDNQITLYKGGPQDRAAPTSTYFDKVEGP 88
QY 55 NETKNVVEIAAIIISATOGYNGTIPAYQTSAGTKTYTMGSEDHGVIPRAIHDFQ 114
Db 89 ASQTEVVEEGAKOVAMSVLSGINATIFAYQTSAGTKTYTMGITEY-----GVTESAVNDIVR 142
QY 115 KIKFPDFEFLRYSMEIYNETITDLCGQTKMKPLIREDVNRNVVADLVEEVYVTS 174
Db 143 HIENTPERDFIILKSAMEIYNEIVKDLL--RPESTNLELDDPEKGTIVEKLEEEIAKDS 200
QY 175 EMALKWITKEKSRHYGETKMNQSRSSHTIFRMILESREKGEPSNCEGSKVYSHNLNLDV 234
Db 201 QHLRLHLSICEQVQVGETALNDSRSHQILRMVTESSRLR-EVSGGVKSF-VANLNFVD 258
QY 235 LAGSERAQTGAAGVRLKEGCNINRSFILQGVTKLSDGQGVGFYINRDSKLTIRLQNS 294
Db 259 LAGSERAQAQTHAVGARLKEGCHINRSLLTTLTVIRKLSDDKRSCHIFPYRDSKLTIRLQNS 318
QY 295 LGGNPKTRIICTIPVVSF--DETTLTALQFASAKYMKNTFYVNEVSTDEALLKRYKEI 351
Db 319 LGGNARTALICTWSPACHIEQSRNTXLFATCAKEVNNAKVMNVVMSDQALVKHLOMEV 378
QY 352 MDLKKQLEEVSLTEAQA--AMEKDLQALLEEKDL--LQKVQN---EKIENLTRL 400
Db 379 ARLEAEALRTDPRASSSSIIIMERDRKIRQVQVEKEMEELKKQORDNAQSKLELQKQM 434

RESULT 6

US-10-425-114-62672
; Sequence 62672, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28

```

118 KFPDREFLLRVSYMEIYNETITDLLCTQKMKPLIIREDVNRNVYVADLTTEEVVYVTSEMA 17
   ||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:|||||
128 KKSDFEFTLVSVYLEIYNEEILDLDCPSREKAQINIREDPKEGKIIVGLTEKTVLVALDT 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
178 LKWITTKGSKSHYGETKNQSSRSHTIFRMILRESREKGSPNCEGSVKVSHLNLVDLAG 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
188 VSCLEQGNNSETVASTAMNSOSSRSHAIFTISLEQRKKSD----KNSSFSLHLVDLAG 243
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
238 SERAAQTGAAGVRUKBEGCINNRSFILGVIKKLSDGVGGFINRYRDKLTRILQNSLGG 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
244 SEROKTVAEGDRLEKGINIRGLCLGNVISALGDCKGSGFVYRDYSKLTRIQDSLGG 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
298 NPKTRIICTITPV--SFDETITALQFASTAKYMNTPYNEVSTDDEALLKYRKIEIMDLX 355
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
304 NGHSLMWIACVSPADSNLBETINTLYRADRAKINKKPIVN-----IDPQTAEMLNHLK 355
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
356 KOLEEVSL-----ETRAQAEMEKQLAOLLEEKOLLQKVONEKIENL 390
   ||:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
356 QOVQOQLVILLQAHGGTLPGSTVEPSENLOSIMEKNQ--SLVENEKLSGLSEAAGQT 413
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
397 TRML 400
   :|||
414 AQML 417

RESULT 8
US-10-408-765A-2153
; Sequence 2153, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale B.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2153
; LENGTH: 1232
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2153

Query March          33.8%; Score 694.5; DB 16; Length 1232;
Best Local Similarity 39.6%; Pred. No. 2.4e-55;
Matches 168; Conservative 66; Mismatches 145; Indels 45; Gaps

      QY      6 VAVCVVRPLRNREEISLG-ETAQVYKTDNNVIYQVDSKSNFDRVFHGNETTKNVYEE 644
      Db      10 VRVALRCRLVPKREISEGCQMCLSFVPGEQQVVGTD--KSFTYDFVFPDPSTEQEVEN 677
      QY      65 IAAPIIDSATQGYNGTTTFANGQTASGKTYTMWG-----SRDHGVIIPRAIHDFOKIK 11
      Db      68 AVAPLIKGVFKGYNATVLAYGQTGSGKTSYMGGAYTAOBENPTVGVIIPRVQLLKFEID 12
      QY      118 KFPDREFLLRVSYMEIYNETITDLLCTQKMKPLIIREDVNRNVYVADLTTEEVVYVTSEMA 17
      Db      128 KKSDFEFTLVSVYLEIYNEEILDLDCPSREKAQINIREDPKEGKIIVGLTEKTVLVALDT 18
      QY      178 LKWITTKGSKSHYGETKNQSSRSHTIFRMILRESREKGSPNCEGSVKVSHLNLVDLAG 23
      Db      188 VSCLEQGNNSETVASTAMNSOSSRSHAIFTISLEQRKKSD----KNSSFSLHLVDLAG 24
      QY      238 SERAAQTGAAGVRUKBEGCINNRSFILGVIKKLSDGVGGFINRYRDKLTRILQNSLGG 2

```


; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 169150
; LENGTH: 965
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_67597C.1.pap
US-10-437-963-169150

Query Match 33.7%; Score 691.5; DB 16; Length 965;
Best Local Similarity 40.8%; Pred. No. 3.2e-55;
Matches 166; Conservative 70; Mismatches 116; Indels 55; Gaps 10;

QY 1 AEEGAVAVCVVRPLNSRRE-----ESLGETAQVYWKTDNNVYQVDSKSFNDRVPH 53
Db 17 AKERIMVSVRLPLNGREAGDSCEISPTVMFRST---VPERAMPEPTAYTVDRVFG 73
QY 54 GNETTKNVYBEIAAPIIDSAIOGYNGTIFAYGOTASGKTYMMGSEDHLGVIPRAIHDF 113
Db 74 PDSSTQVYEEGAKVALSVVSGINGSIIFAYGOTSSGKTYMTGITEY-----SVLDIY 127
QY 114 QKIKKPPDFEFLRVSYMEIYNETITDLCGTQKMKPLIIRREDVNRNVYVADLTEEVTY 173
Db 128 DYIEKHPERFELFSAIETIYNEAVRDLL--SHDTPLRLDDPEKTTVEKLTETELRD 185
QY 174 SEMALKWITKGEKSRHYGETKMNQSRSSHTIFRMILESEKPEPSNCEGSKVSHNLV 233
Db 186 KDLHRLNLAACEAQRQIGETALNETSSRSQHILR-----LNFV 223
QY 234 DLASERAAQTGAAGVRLKEGCNINRSLFILGOVVKLSGQVGGFYNIRDSKLTIRLON 293
Db 224 DLASERASQTASAGVRLKEGSHINRSLITLKGVRQLSKGR--NGHIPYRDSKLTIRLQS 282
QY 294 SLGNPKTRICTITPV--SFDETLTALQFATAKYMKNTPYVNEVSTDEALLKRYKEI 351
Db 283 SLGNARTALICTNSPARSHIEQRNTLLFATCAKEVNTNAQVNVMSDKALVKHLQREL 342
QY 352 MDLKKOLE---EVSLETRAQAM-EKD-----QLAQLEEKOLLQ 386
Db 343 ERLQSEIKFPAPASCTTHAEALREKDAQIKKLEKQLKELMEERDTVK 389

RESULT 12
US-10-311-642-2
; Sequence 2, Application US/10311642
; Publication No. US20040086878A1
; GENERAL INFORMATION:
; APPLICANT: Cytokinetics, Inc.
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: NOVEL MOTOR PROTEINS AND METHODS FOR THEIR
; FILE REFERENCE: 020552-001910US
; CURRENT APPLICATION NUMBER: US/10/311,642
; CURRENT FILING DATE: 2003-09-29
; PRIOR FILING DATE: 09/597,602
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1029
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: HsKif17 amino acid sequence
; FEATURE: HsKif17
; OTHER INFORMATION: Description of Artificial Sequence: HsKif17
US-10-311-642-2

Query Match 33.7%; Score 691.5; DB 16; Length 1029;
Best Local Similarity 41.0%; Pred. No. 3.5e-55;
Matches 171; Conservative 62; Mismatches 139; Indels 45; Gaps 11;
QY 5 AVAVCVVRPLNSRREESLGETAQVYWKTD-----NNVIYQVDSKSFNDRVPHGNE 56
Db 5 AVKVVVRCRPMNQREEL--RCQPVVTVDCAQAQCCIQNPGGADEPPKQFTFDGAYVHDH 62
QY 57 TTKNVYBEIAAPIIDSAIOGYNGTIFAYGOTASGKTYMMGSED---HLGVIPRAIHDF 113
Db 63 VTEQIYNEIAYPLVEGVTEGYNGTIFAYGOTSGSKSFNMQGLPDDPSQGIIPRAEHVF 122
QY 114 QKIKKPPDFEFLRVSYMEIYNETITDLCGTQKMKPLIIRREDVNRNVYVADLTEEVTY 173
Db 123 ESQCAENTKFLVRASYLEIYNEVDRLDGADTKQK--LELKEHPEKGVVYKGLSMETVHS 181
QY 174 SEMALKWITKGEKSRHYGETKMNQSRSSHTIFRMILESR---EKGEPSNCEGSKVSHL 230
Db 182 VAQCEHIMETGWNRSVGVTLNMKDSRSHSIFTISIEMSAVDERG-----KDLHLAGKL 236
QY 231 NLVDLAGSRAAQTGAAGVRLKEGCNINRSLFILGOVVKLSGQVGGFYNIRDSKLTIRI 290
Db 237 NLVDLAGSROSKTGATGERLKEATKINLSLGNVISALVDGRC--KHVPYRDSKLTIRL 295
QY 291 LQNSLGNPKTRICTITPV--SFDETLTALQFATAKYMKNTPYVNEVSTDEALLKRYR 348
Db 296 LQDSLGNKTYLTMVCLSPADNNYDETLSTLRANRAKINRNKPRINEDPKD--ALLREYQ 354
QY 349 KEIMDLKKOLEVSLLETRAQAMEKDLAQLL-----BEKDLQKQVNEKIE 394
Db 355 EEIKKLKAIL-----TQMSFSSLALLSRQPPDPVQVBEKLLPQVPIQHDME 403

RESULT 13
US-10-116-712-664
; Sequence 664, Application US/10116712
; Publication No. US20030194764A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Switzer, Ann
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.568
; CURRENT APPLICATION NUMBER: US/10/116,712
; CURRENT FILING DATE: 2002-04-07
; NUMBER OF SEQ ID NOS: 670
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 664
; LENGTH: 1232
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-116-712-664

Query Match 33.3%; Score 684.5; DB 14; Length 1232;
Best Local Similarity 39.6%; Pred. No. 2.1e-54;
Matches 168; Conservative 66; Mismatches 145; Indels 45; Gaps 9;
QY 6 VAVCVVRPLNSRREESLG-ETAQVYWKTDNNVYQVDSKSFNDRVPHGNETTKNVYEE 64
Db 10 VRVALRCRPLVPKKEISEGCMCLSEVPGEQPVVGTD--KSFTYDFVDPSTQEVEVT 67
QY 65 IAAPIIDSAIOGYNGTIFAYGOTASGKTYMMG-----SEDHIGVIPRAIHDFQKIK 117
Db 68 AVAPLKGFKGNATVLAAYGQTGSKTYSMGAYTAQENBPTGVIPRVQLLFKEID 127
QY 118 KPPDFEFLRVSYMEIYNETITDLCGTQKMKPLIIRREDVNRNVYVADLTEEVTYSEMA 177
Db 128 KKSDFEFTLVSYLEIYNEEIIIDLLCPSREKAIINREDPKGIKIVGTEKTLVVALDT 187
QY 178 LKWTYKGEKSRHYGETKMNQSRSSHTIFRMILESEKPEPSNCEGSKVSHNLVLDLAG 237
Db 188 VSCLEQGNNSRTVASTAMNSQSRSHAFTI---SLEQKKSKDKNSSFR--SKLHLVDLAG 243

```
QY 238 SERAQTGAAGVRLKEGNCINRSLFILGVITKLSGQGVGFNYRDSKLTILQNSLGG 297
Db 244 SERQKTKAEGDRLKEGINRGLCLGNVISALGDDKGGFAPYRDSKLTILQNSLGG 303
QY 298 NPKTRIICITIPV--SFDETLTALQFASHTAKYKNTPVNVSTDEALLKRYRKEIMDLK 355
Db 304 NSHTLMIACVSPADSNLEETILNRYADRAKIKNKPVN-----IDPQTAEHLNHLK 355
QY 356 KQLEEVSL-----ETRAQAMEKDQLAQLLEEKDLLOKVONKIEIENL 396
Db 356 QVQVQQLQVLLQAHGGTLPGSITVPESENQLSLMEKNQ--SLVEENEKLSRGLSEAGQT 413
QY 397 TRML 400
Db 414 AQML 417

RESULT 14
US-10-116-712-669
; Sequence 669, Application US/10116712
; Publication No. US20030194764A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.568
; CURRENT APPLICATION NUMBER: US/10/116,712
; CURRENT FILING DATE: 2002-04-07
; NUMBER OF SEQ ID NOS: 670
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 669
; LENGTH: 1232
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-116-712-669

Query Match 33.3%; Score 684.5; DB 14; Length 1232;
Best Local Similarity 39.6%; Pred. No. 2.1e-54;
Matches 168; Conservative 66; Mismatches 145; Indels 45; Gaps 9;

QY 6 VAVCVVRVPLANSREESLG-ETAQVYWKTDNNVIVQDGSKSFNFRVPHGNETTKNVYEE 64
Db 10 VRVALRCEPLVPKEISECQMLCFVPGEQPVVGTD--KSTYDFVDPSTEQEVEPT 67
QY 65 IAAPIIDSAIQYNGTIPAYGQTASGKTYTWMG-----SEDLGVIPRAIHDFQKIK 117
Db 68 AVAPLLIKGVFKGNATVLAQYQSGKTYSGGAYTAQENEPYGVIPRVLIQLLFKEID 127
QY 118 KFPDPREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVVYADLTVEVVYTGEMA 177
Db 128 KKSDFEFTLVSYLEIYNEETDILDCPSREKAQINIREDPKEGKIVGLTEKTVLVALDT 187
QY 178 LKWTITKEKSRHYGETKMNORSRSHITFRLMLESREKGEPSNCEGSGVKVSHMLVDLAG 237
Db 188 VSCLEQGNNGRTVASTANSSOSSHAFTI--SLQGGKSKDNSSFR-SKLHLVDLAG 243
QY 238 SERAQTGAAGVRLKEGNCINRSLFILGVITKLSGQGVGFNYRDSKLTILQNSLGG 297
Db 244 SERQKTKAEGDRLKEGINRGLCLGNVISALGDDKGGFAPYRDSKLTILQNSLGG 303
QY 298 NPKTRIICITIPV--SFDETLTALQFASHTAKYKNTPVNVSTDEALLKRYRKEIMDLK 355
Db 304 NSHTLMIACVSPADSNLEETILNRYADRAKIKNKPVN-----IDPQTAEHLNHLK 355
QY 356 KQLEEVSL-----ETRAQAMEKDQLAQLLEEKDLLOKVONKIEIENL 396
Db 356 QVQVQQLQVLLQAHGGTLPGSITVPESENQLSLMEKNQ--SLVEENEKLSRGLSEAGQT 413
QY 397 TRML 400
Db 414 AQML 417
```

```
RESULT 15
US-10-408-765A-1664
; Sequence 1664, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1664
; LENGTH: 672
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1664

Query Match 33.2%; Score 682.5; DB 16; Length 672;
Best Local Similarity 47.3%; Pred. No. 1.3e-54;
Matches 158; Conservative 55; Mismatches 106; Indels 15; Gaps 8;

QY 37 IYQVDGS----KSFNFRVPHGNETTKNVYEBIAPIIDSAIQYNGTIPAYGQTASGKT 92
Db 18 VHKTDSSNEPKTFTFTVFGPESKQLDVNLTAAPIIDSVLEGYNGTIPAYGQTGTGT 77
QY 93 YTMGSE---DHLGVIPRAIHDFQKIKFEP-DREFLLRVSYMEIYNETITDLCGTQKMK 148
Db 78 FTMGVRAIPELRLGIIPNSFAHIFGHIAKABGDTFLRVSYLEIYNEEVRDLL-GKQQT 136
QY 149 KPLIREDVNRNVYADLTVEVVYTGEMALKWTITKEKSRHYGETKMNORSRSHITFIRM 208
Db 137 QRLVEXKRPDVGVIKDLISAVVNNADMDRIMTLGHKRSVGVATNMNHSRSHAFTI 196
QY 209 ILESREKGEPSNCEGSGVKVSHNLVDLAGSERAAQTGAAGVRLKEGNCINRSLFILQVI 268
Db 197 TIRCEKGIQGNMH--VRMGKHLVDLAGSERQAKTGATQRLKEATKINLSLTIGNVI 254
QY 269 KKLSDGQGVGFNYRDSKLTILQNSLGGNPKTRIICITIPV--SFDETLTALQFASHTAK 326
Db 255 SALVDGK-STHVPYRNSKLTILQNSLGGNSKTMNCANIGADYNYDETISTLYANRAK 313
QY 327 YMKNTPVNVSTDEALLKRYRKEIMDLKQLEE 360
Db 314 NIKNKARINEDPKD-ALLRQFQKEIEBELKKLEE 346

Search completed: July 29, 2004, 10:06:27
Job time : 22.1784 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:28:21 ; Search time 6.03629 Seconds
(without alignments)
6437.961 Million cell updates/sec

Title: US-10-045-631b-88_COPY_2_405
Perfect score: 2053
Sequence: 1 AEEGAVAVCVVRPLNSREE.....LQKVQNEKIENLRLMTSS 404

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.78.*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2053	100.0	2663	1 S28261	centromere protein
2	1468.5	71.5	2954	2 T14156	kinesin-related pr
3	872	42.5	823	2 T52425	kinesin-like prote
4	836.5	40.7	888	2 D96619	protein T30E16.9
5	754	36.7	1459	2 T30196	kinesin motor prot
6	723	35.2	701	1 B44259	kinesin-related pr
7	713	34.7	747	1 A57107	kinesin-related pr
8	712.5	34.7	742	1 S58691	kinesin-related pr
9	708.5	34.5	786	2 A53939	kinesin homolog KH
10	702	34.2	699	1 S38982	kinesin-related pr
11	694	33.8	1121	2 T06065	kinesin-related pr
12	693	33.8	1231	2 A54803	microtubule-associ
13	692	33.7	909	2 H86350	hypothetical prote
14	687	33.5	1226	2 T51617	hypothetical prote
15	686.5	33.4	581	2 F84599	kinesin-like prote
16	683	33.3	968	2 T45746	probable kinesin h
17	678	33.0	1130	2 T21134	hypothetical prote
18	677	33.0	932	2 T49235	hypothetical prote
19	667	32.5	1225	2 A56514	kinesin-like prote
20	662.5	32.3	1066	1 A48669	chromokinesin - ch
21	659.5	32.1	1263	2 T13465	kinesin-related pr
22	656	32.0	672	2 S54351	hypothetical prote
23	652	31.8	975	1 A31497	kinesin osm-3 - Ca
24	651	31.7	1254	2 T18277	kinesin heavy chai
25	649	31.6	1921	2 T13827	kinesin-73 - fruit
26	648.5	31.6	793	2 JCS831	kinesin-related pr
27	647.5	31.5	834	2 T06055	hypothetical prote
28	645	31.4	1031	1 A38713	kinesin heavy chai
29	644	31.4	1076	2 B84687	probable kinesin-l

30	637	31.0	967	1 A35075	kinesin heavy chai
31	634.5	30.9	1032	2 I38510	neuronal kinesin h
32	629.5	30.7	963	1 A41919	kinesin heavy chai
33	627	30.5	1027	2 S37711	kinesin heavy chai
34	625	30.4	784	1 A55236	kinesin-related pr
35	622.5	30.3	1388	2 T30335	KLP2 protein - Afr
36	622	30.3	843	2 S44868	kinesin heavy chai
37	622	30.3	1576	2 T29237	hypothetical prote
38	622	30.3	1584	1 JN0114	kinesin-related pr
39	622	30.3	1584	2 T15822	kinesin-like prote
40	621	30.2	294	2 S38983	kinesin-related pr
41	620	30.2	554	2 T50118	kinesin-related pr
42	620	30.2	857	2 E84600	probable kinesin h
43	616	30.0	968	2 T51933	kinesin motor prot
44	613	29.9	1022	2 E84792	probable kinesin h
45	610.5	29.7	881	2 I84737	kinesin heavy chai

ALIGNMENTS

RESULT 1

S28261
centromere protein E - human
N:Alternate names: centromere 312K protein; kinesin-related protein CENP-E
C:Species: Homo sapiens (man)
C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C:Accession: S28261
R:Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
Nature 359, 536-539, 1992
A:Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.
A:Reference number: S28261; MUID:93024922; PMID:1406971
A:Accession: S28261
A:Molecule type: mRNA
A:Residues: 1-2663 <YEN>
A:Cross-references: EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865
C:Genetics:
A:Gene: GDB:CENPE
A:Cross-references: GDB:361164; OMIM:117143
A:Map position: 4q24-q25
C:Superfamily: centromere protein E; kinesin motor domain homology
C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
F:7-335/Domain: kinesin motor domain homology <KMT>
F:86-93/Region: nucleotide-binding motif A (P-loop)
F:486-2183/Domain: coiled coil #status predicted <COI>
F:92/Binding site: ATP (lys) #status predicted

Query Match	100.0%;	Score 2053;	DB 1;	Length 2663;
Best Local Similarity	100.0%;	Pred. No. 4.5e-125;		
Matches 404;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	AEEGAVAVCVVRPLNSRESLGETAQVYWKTDNNVIYQVDGSKSFNDFRVFHNQTKN	60	
Db	2	AEEGAVAVCVVRPLNSRESLGETAQVYWKTDNNVIYQVDGSKSFNDFRVFHNQTKN	61	
QY	61	VVEETAAPIIDSALQVNGTIFAYGQTASGKTYTMGSEDHLGVIPRAIHDFQKIKKFP	120	
Db	62	VVEETAAPIIDSALQVNGTIFAYGQTASGKTYTMGSEDHLGVIPRAIHDFQKIKKFP	121	
QY	121	DREFLLRVSMYEIYNETITDLCGTQKWKPLIIRDVNRNVYVADLTVEWVYVTSMAKWK	180	
Db	122	DREFLLRVSMYEIYNETITDLCGTQKWKPLIIRDVNRNVYVADLTVEWVYVTSMAKWK	181	
QY	181	ITKGEKSRHYGETKMNQRSSSHITFRMILSRKGEPSNCEGSKVSVSHNLVDLAGSR	240	
Db	182	ITKGEKSRHYGETKMNQRSSSHITFRMILSRKGEPSNCEGSKVSVSHNLVDLAGSR	241	
QY	241	AAQTGAAGVRLKEGCNINRSLFILGVIKKLSDGQVGGFINVRDSKLTIRLQNSLGGNPK	300	
Db	242	AAQTGAAGVRLKEGCNINRSLFILGVIKKLSDGQVGGFINVRDSKLTIRLQNSLGGNPK	301	
QY	301	TRIICTITPVSFDETLTALQFASHTAKYMKNTPVYNEVSTDEALLKRYRKEIMDLKKQLEE	360	

Db 302 TRIICITPVSFDETLTALQFASAKYMNKTPYVNEVSTDEALLKRYRKEIMDLKKOLEE 361

QY 361 VSLFTRQAMKQDLAQLLEKDLLOKQVNEKIENLTMLVTSS 404

Db 362 VSLFTRQAMKQDLAQLLEKDLLOKQVNEKIENLTMLVTSS 405

RESULT 2

T14156

kinesin-related protein - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000

C:Accession: T14156

R:Wood, K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.

Cell 91, 357-366, 1997

A:Title: CENP-E is a plus end-directed kinetochore motor required for metaphase chromosome segregation

A:Reference number: Z17893; PMID:98028574; PMID:9363944

A:Accession: T14156

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2954 <WOO>

A:CROSS-references: EMBL:AF027728; NID:g2586070; PID:g2586071; PIDN:AAC60300.1

C:Genetics:

A:Gene: XENP-E

C:Superfamily: centromere protein E; kinesin motor domain homology

Query Match 71.5%; Score 1468.5; DB 2; Length 2954;

Best Local Similarity 70.9%; Pred. No. 5.4e-87;

Matches 290; Conservative 52; Mismatches 60; Indels 7; Gaps 3;

QY 1 AEGGAVACVVRPLNSRESLGETAQVYKTDNNVIYQVDSKSNFDFVHNETTKN 60

Db 2 SEGDAVACVVRPLNQEQ--GDQANLQWKAGNTISQVDTGKSNFDFVHNETTKN 59

QY 61 VYEEIAAPLIDSAIQVNGTIFAYGQASGKTITMGSSEHGLGVIPRAIHDFQKIKFP 120

Db 60 IYGEIAVPIIRSAIQVNGTIFAYGQASGKTITMGSSEHGLGVIPRAIHDFQKIKFP 119

QY 121 DRPEFLVSYMEIYNEITDILGCTQKMKPLIREDVNRNVYADLTVEEVVYVSEMAKW 180

Db 120 NREFLLVSYMEIYNEITDILGCTQKMKPLIREDVNRNVYADLTVEEVVYVSEMAKW 179

QY 181 ITGKGRHRYGVTQKQSSRSHTIFRMILSRKSGPS--NCEGSKVYSHLNLVDLAG 237

Db 180 IKGKGRHRYGVTQKQSSRSHTIFRMILSRKSGPS--NCEGSKVYSHLNLVDLAG 239

QY 238 SERAATCAAGVRLKEGNCINRSFILGQVILKSLDQGVGFYNYRDSKLTRELQNSLGG 297

Db 240 SERASQTGAEGVRLKEGNCINRSFILGQVILKSLDQGVGFYNYRDSKLTRELQNSLGG 299

QY 298 NPKTRITITPVSFDETLTALQFASAKYMNKTPYVNEVSTDEALLKRYRKEIMDLKKQ 357

Db 300 NAKTVIITITPVSFDETLTALQFASAKYMNKTPYVNEVSTDEALLKRYRKEIMDLKKQ 359

QY 358 LE--EVSLETRAQAMEKQDLAQLLEKDLLOKQVNEKIENLTMLVTSS 404

Db 360 LENLESSETRAQAMEKQDLAQLLEKDLLOKQVNEKIENLTMLVTSS 408

RESULT 3

T52425

kinesin-like protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000

C:Accession: T52425

R:Kato, A.; Suzuki, M.; Kuwahara, A.; Ooe, H.; Higano-Inaba, K.; Komeda, Y.

Gene 239, 309-316, 1999

A:Title: Isolation and analysis of cDNA within a 300 kb Arabidopsis thaliana genomic region

A:Reference number: Z25171

A:Accession: T52425

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-823 <KAT>

A:CROSS-references: EMBL:AB028468; PIDN:BAA88112.1

C:Genetics:

A:Gene: ZCF125

Query Match 42.5%; Score 872; DB 2; Length 823;

Best Local Similarity 47.6%; Pred. No. 6.1e-49;

Matches 204; Conservative 58; Mismatches 123; Indels 44; Gaps 9;

QY 6 VAVCVVRPLNSRESLGETAQVYKTDNNVIYQVDSKSNF--FDRVPHGNE 56

Db 4 ICVAVVRP-----PAPENGASLWKVEDN---RISLHKSLDPTTTTASHAFDHFDESS 54

QY 57 TTKNVYEEIAAPLIDSAIQVNGTIFAYGQASGKTITMGSSEHGLGVIPRAIHDFQK 116

Db 55 TNASVYELTKDIIHAAVEGFGTAFAYGQTSSTGKFTWTGSETDPTGIILRRSVDRVFERI 114

QY 117 KKFDPREFILRYSYMEIYNEITDILGCTQKMKPLIREDVNRNVYADLTVEEVVYVSEM 176

Db 115 HMSDREFILRYSYMEIYNEEINDLL--AVENQRLQIHEHLRGVFAVLKEEIVSDAQ 172

QY 177 ALKWTGKSKSHRYGVTQKQSSRSHTIFRMILSRKSGPSNCEGSKVYSHLNLVDLA 236

Db 173 ILKLDSGEVNRHFGTNNVHSSRSHTIFRWVIESR--GKNSSSDAIRVSVNLVDLA 230

QY 237 GSERAATCAAGVRLKEGNCINRSFILGQVILKSLDQGVGFYNYRDSKLTRELQNSL 295

Db 231 GSERIAKTGAGVRLQEGKYINKSLMILGNVINKLSDSTKLRAHIPYRDSKLTRELQNSL 290

QY 296 GGNPKTRITITPVSV--FDETLTALQFASAKYMNKTPYVNEVSTDEALLKRYRKEIMD 353

Db 291 GGNAKTCTIITTAPEEHIEESKGTQFASRAKRITNCAQVNEILDAALLKQKLEIEE 350

QY 354 LKQLEEVSLFTRAQ-----AMEKQDLAQLLEKDLLOKQVNE-----KIE 394

Db 351 LEMKLGSHAEVLEQILNLSNMLKVECEKLTQLEBEKQKQEQENCIEKQEQMKIE 410

QY 395 NLTRMLVTS 403

Db 411 NLNFTVNS 419

RESULT 4

D96619

protein T30E16.9 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: D96619

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; PMID:21016719; PMID:11130712

A:Accession: D96619

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-888 <STO>

A:CROSS-references: GB:AE005173; NID:g8778739; PIDN:AAF79747.1; GSPDB:GN00141

C:Genetics:

A:Gene: T30E16.9

A:Map position: 1

Query Match 40.7%; Score 836.5; DB 2; Length 888;

Best Local Similarity 43.5%; Pred. No. 1.4e-46;

Matches 204; Conservative 58; Mismatches 126; Indels 81; Gaps 10;

QY 6 VAVCVVRPLNSRESLGETAQVYKTDNNVI-----Y 38

```

Db 4 ICVAVKVRP-----PAPENGASLWKVEDNRISLHKSLDTPITTAASHAFVSGISISTDLI 57
QY 39 QVDSKSNF-----DRVEHGNETTKNVYBEIAAPIIDSALQGYNGTIFAYGOTA 88
Db 58 EIVSLFLSFGVVYFFLLADHVDESSNVAIVYELLTKDIIHAAVEGNGTAFAYGOTS 117
QY 89 SKTYTMMGSDHLGVIPRAIHDI FQIKKFPDREFLLRVSYMEIYNETITDLLCGTQRM 148
Db 118 SKTFTMTGSETDPIIRSRVDRVFERIHWISDREFLLRVSYMEIYNEINDDL--AVEN 175
QY 149 KPLIREDVNRNVYADLTVEVYIYSEMALKWITKGEKSRHYGETKQNRSSRSHITFRM 208
Db 176 QRLQIHEHLERGFFVAGLKEEIVSDAQILKIDSGEVNRHFGETMNVHSSRSHITFRM 235
QY 209 -----ILSREKGEPSNCGSVKVSGLNLVDLAGSRAAQTGAAGVRLEKGCN 256
Db 236 VFRSYERDLLLVIESR--GKNSSSDAIRVSVNLVDLAGSERIAKTGAGVRVLEQGY 293
QY 257 INRSFLIQQVVKLSG--QVGGFINYRDSKLTILQNSLGGNPKTRIICTITPVS--PD 313
Db 294 INKSLMILGNVINKLSDSSTKRAHIPYRDSKLTILQPALGNAKTCIICTIAPREEHIE 353
QY 314 EFLTALQFASAKYMNTPYNEVSTDEALLKRYKREIMDLKQLEVSLETRAQ----- 368
Db 354 ESKGTLOFASRAKRTNCAQVNEIILTDAAALKRQKLEIPELRMKLQGSFARVLEQBI 413
QY 369 -----AMEKDQLAQLLEEKDLOKQVNE-----KIENLTRMLVTS 403
Db 414 SNQMLKYLECEBLTKQLEEBKQKEQENCICEQOMKIENLNNTNS 462

RESULT 5
T30196
Kinesin motor protein 1 - smut fungus (Ustilago maydis)
C:Species: Ustilago maydis (corn smut)
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C:Accession: T30196
EMBO J. 16, 3464-3473, 1997
R:Lehmler, C.; Steinberg, G.; Snettselaar, K.M.; Schliwa, M.; Kahmann, R.; Bolker, M.
A:Title: Identification of a motor protein required for filamentous growth in Ustilago
A:Reference number: Z20770; MUID:97361828; PMID:9218789
A:Accession: T30196
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1459 <LEH>
A:Cross-references: EMBL:U92844; NID:g2062749; PID:g2062750; PIDN:AAB63336.1
C:Genetics:
A:Gene: kin1
C:Function:
A:Description: required for filamentous growth in Ustilago maydis

Query Match 36.7%; Score 754; DB 2; Length 1459;
Best Local Similarity 38.2%; Pred. No. 6.2e-41;
Matches 190; Conservative 70; Mismatches 127; Indels 110; Gaps 14;

QY 6 VAVCVVRPLNSREESLGETAQYVWKTDN--NVIIYQVD----- 41
Db 237 VVVCVRMRP--SRASSSDSEASV-WNCDSKRNRIPTTEHHPALAKRTTSSERAGAGASIA 293
QY 42 -----GSKSFNPDVFNHNETTKNVYBEIAAPIIDSALQGYNGTIFAYGQ 86
Db 294 AAPSSHDLDHEDPTSTYHFQDKLITGAQTDDMYHSHIAPVRAAIVEGYNGTIFAYGQ 353
QY 87 TASGKTYTMMGSEDHLGVIPRAIHDI FQIKKFPDREFLLRVSYMEIYNETITDLLCGTQ 146
Db 354 TGSCKTHTMSGSDAEFGVIPRAVEGIFQMIKDEPDREFLLRVSYLEIYNETILKOLLA--- 410
QY 147 KMKPL-----IREDVNRNVVADLTVEV 171
Db 411 PLPPLTGTSGSLQTTDRPASPIKGGSSHAQSQSCITRIIEDQKSRVITGLUREIV 470
QY 172 YTEMALKWKTKGSRHYGETKQNRSSRSHITFRMILSREKGEPSNCGSVKVSGLN 231

```

```

Db 471 TDANTVLCIORQDQDERHVGAIDWNERSRSHCVFQLTIESRSPASASKE--VRISQLN 528
QY 232 IYDLAGSBAATGAGVPLKGCNINRSLFILGQVVKLSGQVGG--FINYRDSKLT 289
Db 529 LIDLAGSERAA---SQAERKEGAFINKSLTLGTIVIGKLTPEVNGDAHPIYRDSKLT 585
QY 290 ILQNSLGGNPKTRIICTITPVS--PDFTLALQFASAKYMNKTPVYNEVSTDEALLKRY 347
Db 586 ILQTSLSGNARLAVICTISPDTEHANETLSTLKFGRCKLVVTTAKKGVTAMDKALLQKY 645
QY 348 RKEIMDLKKQLEB-----VSLETRAQAMEK-DQLAQLLE-----EKDLLQKVQ 389
Db 646 RKELDALARAKLANGSPNPNVMTIVSAESKESQKLDQLNQKEAAQREVEDNQKR 705
QY 390 NE---KIENLTRMLVTS 403
Db 706 SHLKAQIEHLTRLILTS 722

RESULT 6
B44259
Kinesin-related protein KIF3A - mouse
C:Species: Mus musculus (house mouse)
C>Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C:Accession: B44259; S27872
R:Aizawa, H.; Sekine, Y.; Takemura, R.; Zhang, Z.; Nangaku, M.; Hirokawa, N.
J. Cell Biol. 119, 1287-1296, 1992
A:Title: Kinesin family in murine central nervous system.
A:Reference number: A44259; MUID:93077686; PMID:1447303
A:Accession: B44259
A:Molecule type: mRNA
A:Residues: 1-701 <AIZ>
A:Cross-references: EMBL:D12645; NID:g220469; PID:BAA02166.1; PID:g220470
A:Experimental source: Brain
A:Note: sequence extracted from NCBI backbone (NCBIP:118911)
C:Complex: heterodimer with KIF3B (PIR:A57107); the KIF3A/3B heterodimer associates with
C:Function:
A:Description: KIF3 complex is a motor protein that provides anterograde fast axonal tra
A:Superfamily: kinesin-related protein KIF3; kinesin motor domain homology
C:Keywords: ATP; coiled coil; heterodimer; heterotrimer; microtubule binding; nucleotide
F:1-368/Domain: head globular #status predicted <HGL>
F:15-351/Domain: kinesin motor domain homology <KMOT>
F:100-107/Region: nucleotide-binding motif A (p-loop)
F:369-595/Domain: helical rod #status predicted <ROD>
F:600-701/Domain: tail globular #status predicted <TGL>
F:106/Binding site: ATP (lys) #status predicted

Query Match 35.2%; Score 723; DB 1; Length 701;
Best Local Similarity 44.4%; Pred. No. 2.4e-39;
Matches 178; Conservative 66; Mismatches 137; Indels 20; Gaps 10;

QY 6 VAVCVVRPLNSREESL--GETAQYVWKTNNVIYQVDGS----KSFNPDVFNHNETTK 59
Db 15 VVVVRCPLNREKSMCYQAVSDVDEMRTITVHKTDSSNEPKTFTFTDVFVGESQL 74
QY 60 NVYBEIAAPIIDSALQGYNGTIFAYGOTASGTYTMMGSE---DHLGVIPRAIHDI FQKI 116
Db 75 DVYNLTARPIIDSLEGGNGTIFAYGQGTGKTFTMEGVRAPGLRGVIPNSFAHIFGHI 134
QY 117 KKFP--DREFLLRVSYMEIYNETITDLLCGTQKMKPLIREDVNRNVVADLTVEVYVTS 175
Db 135 AKAESGDTFLVRVSYLEIYNEEVRDLL--GKQOTQRLVKERPDGVYIKDLSAYVNNAD 193
QY 176 MALKWITKGEKSRHYGETKQNRSSRSHITFRMILSREKGEPSNCGSVKVSGLNVDL 235
Db 194 DMDRIMTLGHKNRGVGTANNNEHSSRSHAITTITIECKEKGVDGNMH--VRMGKHLVDL 251
QY 236 AGSRAAQTGAAGVRLKEGCNINRSLFILGQVVKLSGQVGGFINYRDSKLTILQNSL 295
Db 252 AGSRQAKTGAQTGQRLKEATKINLSLTGLNVISALVDGK--STHVPYNSKLTILQDLS 310
QY 296 GGNPKTRIICTITPV--SFDFTLALQFASAKYMNKTPVYNEVSTDEALLKRYKREIMD 353

```

Db 311 GGNKTMWCANIGPADNYDYEITSTLYRANRAKNIKNKARINEDPKD-ALLRQFQKEIEE 369
 QY 354 LKQQL---BEVSLTETRAQAMEKDQLAQLLEKDLLOKQVNE 391
 Db 370 LKKKLEGEVSGSDISGEEDDEGELGDEGKRRKRDQ 410

RESULT 7

A57107
 Kinesin-related protein KIF3B - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change 19-Jan-2001
 C:Accession: A57107
 R;Yamazaki, H.; Nakata, T.; Okada, Y.; Hirokawa, N.
 J. Cell Biol. 130, 1387-1399, 1995
 A:Title: KIF3A/B: a heterodimeric kinesin superfamily protein that works as a microtubule
 A:Reference number: A57107; PMID:96032268; PMID:7559760
 A:Accession: A57107
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-747 <YAM>
 A:Cross-references: GB:D26077; NID:g1060922; PIDN:BAA05070.1; PID:g1060923
 A:Experimental source: Brain
 C:Complex: heterodimer with KIF3A (PIR:B44259); the KIF3A/3B heterodimer associates with
 C:Function:
 A:Description: KIF3 complex is a motor protein that provides anterograde fast axonal tra
 C:Superfamily: kinesin-related protein Kif3; kinesin motor domain homology
 C:Keywords: ATP; coiled coil; heterodimer; heterotrimer; microtubule binding; nucleotide
 F:1-363/Domain: head globular #status predicted <HGL>
 F:10-346/Domain: kinesin motor domain homology <KMT>
 F:96-103/Region: nucleotide-binding motif A (P-loop)
 F:364-592/Domain: helical rod #status predicted <ROD>
 F:594-747/Domain: tail globular #status predicted <TGL>
 F:102/Binding site: ATP (Lys) #status predicted

Query Match 34.7%; Score 713; DB 1; Length 747;
 Best Local Similarity 41.4%; Pred. No. 1.2e-38;
 Matches 177; Conservative 69; Mismatches 144; Indels 38; Gaps 10;
 QY 5 AVAVCVVRPLNSREESLG-----ETAAQVYKTDNNVIYQVDSKSPNFRVPHGN 55
 Db 9 SVRVVRCRPNMGKEKAASYDKVDVVLQGVSVKPKGTSHEM--PKTTFDAVYDWN 66
 QY 56 ETTKNVVEETAAPIIDSAIQYNGTIFAYGQTASGKTYTM---MGSEDLGVIPRAIHDI 112
 Db 67 AKQELDYDETRPLVDVSLQGFNGTIFAYGQTGKTVMGVRGDPKRGVIPSFDHI 126
 QY 113 POKTKPPDRFLRVSYMEIYNETITDLCGQKMKPLIREDVNRVYVADLTVEVY 172
 Db 127 FTHISRSNQYLVRASYLETIYQBEIRDLSKDQ--TKLELKERPDGTGVYVKDLSFFYTK 185
 QY 173 TSEMAKWIYTKGSRHYGTEKMNORSRSHITFRMILESEKGEPSNCGSVKVSHLNL 232
 Db 186 SVKEIEHVMVNGNVRSGVATNMHSSRSHAFVITIECEVG--LDGENHVRGKLN 243
 QY 233 VDLASERAQTGAAGVRLKEGCNINRSLFLGQVVKLSGQVGGFINYRDSKLTRELQ 292
 Db 244 VDLASGEQAKTGAGGERLKEATKINLSALGNVISALVDGK-STHIPYRDSKLTRELQ 302
 QY 293 NSLGNPKTRIICTITPVS--DETTLAQFASAKYKNTPYNEVSTDEALLKRYKE 350
 Db 303 DSLGNAKTVMVANMGPSYNFDETTTLRYANRAKNIKNKPRVNEPKD-ALLREFQEE 361
 QY 351 IWDLKKQLREVSLETRAQ-----AMEKDLQAQLLEE-----KDLQKQVNEKI 393
 Db 362 IARLAQAQLKEKISGRKREKREKREGGSGGGGEEEGEEGEEEDGDKDDYWRQOEKL 421
 QY 394 ENLITRLV 401
 Db 422 EIEKRAIV 429

RESULT 8

S58691

kinesin-related protein KRP95 - sea urchin (Strongylocentrotus droebachiensis)
 N;Alternate names: kinesin-2 chain B; KRP (85/95) 95K chain
 C:Species: Strongylocentrotus droebachiensis
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
 C:Accession: S58691
 R;Rashid, D.J.; Wedaman, K.P.; Scholey, J.M.
 J. Mol. Biol. 252, 157-162, 1995
 A:Title: Heterodimerization of the two motor subunits of the heterotrimeric kinesin, KRP
 A:Reference number: S58691; PMID:95404610; PMID:7674298
 A:Accession: S58691
 A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
 A:Molecule type: mRNA
 A:Residues: 1-742 <RAS>
 C:Complex: heterotrimer of a 115K chain and two kinesin-related chains of 85K (PIR:S5899
 C:Superfamily: kinesin-related protein Kif3; kinesin motor domain homology
 C:Keywords: ATP; heterotrimer; microtubule binding; nucleotide binding; P-loop
 F:9-345/Domain: kinesin motor domain homology <KMT>
 F:95-102/Region: nucleotide-binding motif A (P-loop)
 F:101/Binding site: ATP (Lys) #status predicted

Query Match 34.7%; Score 712.5; DB 1; Length 742;
 Best Local Similarity 41.3%; Pred. No. 1.3e-38;
 Matches 175; Conservative 65; Mismatches 137; Indels 47; Gaps 10;
 QY 6 VAVCVVRPLNSREESLGETAQVYKTDNNVIYQV-----DGSKSFNFRVPHGNETT 58
 Db 9 VKVVRRCRPNMSKEISQGHKRIVEMDNKRGLEVNTNPKGPPGEPNKSFTTDTVYDWNKQ 68
 QY 59 KNYVEETAAPIIDSAIQYNGTIFAYGQTASGKTYTMG---SEDHLGVIPRAIHDIPOK 115
 Db 69 IDLYDETRFLRVSYMEIYNETITDLCGQKMKPLIREDVNRVYVADLTVEVY 128
 QY 116 IKKPPDRFLRVSYMEIYNETITDLCGQKMKPLIREDVNRVYVADLTVEVY 175
 Db 129 IARTNQQLVRASYLETIYQBEIRDLLAKDQ--KRLDLKERPDGTGVYVKDLSFFYTKSVK 187
 QY 176 MALKWIYTKGSRHYGTEKMNORSRSHITFRMILESEKGEPSNCGSVKVSHLNLVDL 235
 Db 188 EIEHVMVTVGNRNSVGTNNMHSRSHAFIITIECELGVDG--ENHIVRGKLNVDL 245
 QY 236 AGSERAQTGAAGVRLKEGCNINRSLFLGQVVKLSGQVGGFINYRDSKLTRELQ 295
 Db 246 AGSERQAKTGATGDLKEATKINLSALGNVISALVDGK--SSHIPYRDSKLTRELQ 304
 QY 296 GGNPKTRIICTITPVS--DETTLAQFASAKYKNTPYNEVSTDEALLKRYKEIMD 353
 Db 305 GGNKTVMVANMGPSYNFDETTTLRYANRAKNIKNKPKINEDPKD-ALLREFQEEISR 363
 QY 354 LKQQL-----EEVSLTETRAQAMEKDQLAQLLEE-----KDLQKQV 388
 Db 364 LKQALDKGPSDGRKKKPKGQGGDDDDIEDTEBEGDEMDE-----EEMYKESQQL 418
 QY 389 QNEK 392
 Db 419 EEEK 422

RESULT 9

A53939
 kinesin homolog Khp1 - Chlamydomonas reinhardtii
 C:Species: Chlamydomonas reinhardtii
 C:Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 02-Feb-2001
 R;Walther, Z.; Vaeshishta, M.; Hall, J.L.
 J. Cell Biol. 126, 175-188, 1994
 A:Title: The Chlamydomonas FLA10 gene encodes a novel kinesin-homologous protein.
 A:Reference number: A53939; PMID:94299638; PMID:8027176
 A:Accession: A53939
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-786 <WAL>
 A:Cross-references: EMBL:L33697; NID:g497696; PIDN:AAA21738.1; PID:g497697

A;Note: authors translated the codon AAC for residue 753 as Asp

C;Genetics:

A;Gene: FLA10

C;Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology

C;Keywords: ATP; coiled coil; nucleotide binding; P-loop

F;11-359/Domain: kinesin motor domain homology <KMOT>

F;97-104/Region: nucleotide-binding motif A (P-loop)

Query Match 34.5%; Score 708.5; DB 2; Length 786;

Best Local Similarity 44.4%; Pred. No. 2.5e-38;

Matches 167; Conservative 66; Mismatches 118; Indels 25; Gaps 9;

QY 5 AVAVCVRVRLNREESLSIGTAQVYVKTNNVLY-----QVDGS---KSENFDRVFGNET 57

DB 10 SVKVVRCRPLNGKEKADGRSIRVDMVDAGQVKNRPKADASEPPKAFITFDQVYDNQC 69

QY 58 TKNVYEIAAIIIDSAIQGYNGTIFAYGQTASGKTYTMMGSEDH---LGVIPRAIHDFQ 114

DB 70 QRDVFDITAPRLIDSCIEGYNGTIFAYGQTGKSHTWEGKDEPPELRLGLIPNTRYVFE 129

QY 115 KI-KKFPDRFLRVSMYINETITDLCGTQKMKPLIREDVNRNVYVADLVEEYVT 173

DB 130 IARDSGTKEFLVRSSYLEIYNEVRDL--GKDSKMKELKESPDGRGVYVKDLSQVCKN 188

QY 174 SEMALKWITKGEKSRHYGETKMQRSRSHITFIRMILESREK-----GEPSNCEGS 224

DB 189 YEEMKVKLLAGKNRQVATLMNQDSSRSRSHITFIEIEKLSAAAKPKGAKKDSNH 248

QY 225 VKVSHLNLVDLAGSERAQAOTGAAGVRKKEGNCINRSFILGOVVKKLSGQGVGFINRYD 284

DB 249 VRVGNLNLVDLAGSERQDKTGATGDLKEGKIKNLSTALGNVISALVDGK-SGHIPYRD 307

QY 285 SKLFRILQNSLGNPKTRIICTITPV--SPDETALQFATAKYMKNTPYVNEVSTDEA 342

DB 308 SKLRLQLDSLGNGTKTVMWANGPADWNYDETSTLYANRAKNIQNKPKINEDPKD-A 366

QY 343 LKRYRKEIMDLKKQL 358

DB 367 MLRQFQEEIKLKBQL 382

RESULT 10

S38982

Kinesin-related protein KRP85 - sea urchin (Strongylocentrotus purpuratus)

N;Alternate names: kinesin-2 chain A; KRP (85/95) 85k chain

C;Species: Strongylocentrotus purpuratus (purple urchin)

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001

C;Accession: S38982

R;Cole, D.G.; Chinn, S.W.; Wedaman, K.P.; Hall, K.; Vuong, T.; Scholley, J.M.

Nature 366, 268-270, 1993

A;Title: Novel heterotrimeric kinesin-related protein purified from sea urchin eggs.

A;Reference number: S38982; MUID:94050179; PMID:8232586

A;Accession: S38982

A;Molecule type: mRNA

A;Residues: 1-699 <COL1>

A;Cross-references: EMBL:U16993; NID:G295245; PIDN:AAA16098.1; PID:G295246

A;Accession: S72551

A;Molecule type: protein

A;Residues: 2-5,'X',7-11,59-64,125-132,222-226,'X',228-230 <COL2>

C;Complex: heterotrimer of a 115k chain and two kinesin-related chains of 95k (PIR:S5869

C;Superfamily: kinesin-related protein KIF3; kinesin motor domain homology

C;Keywords: ATP; heterotrimer; microtubule binding; nucleotide binding; P-loop

F;11-348/Domain: kinesin motor domain homology <KMOT>

F;97-104/Region: nucleotide-binding motif A (P-loop)

F;103/Binding site: ATP (Lys) #status predicted

Query Match

Best Local Similarity 34.2%; Score 702; DB 1; Length 699;

Matches 190; Conservative 65; Mismatches 131; Indels 92; Gaps 16;

QY 6 VAVCVRVRLNREESLSIG-----ETAQVYVKTNNVYQVDGSKSFNDRVFHG 54

DB 11 VRVVRVRLNREESLSIG-----ETAPNA-PSGEPPKSTFDIVFAP 66

QY 55 NETTKNVYEIAAIIIDSAIQGYNGTIFAYGQTASGKTYTMMG--SEDLH-GVIPRAIHD 111

DB 67 GAKQTDVYNTARPIVDALIEGYNGTIFAYGQTGKTFTMEGVRSQPELRLGIPNSFAH 126

QY 112 IFQIKKFPDR-BFLRVSMYINETITDLCGTQKMKPLIREDVNRNVYVADLVEEY 170

DB 127 IFGHIAKEQENRFLRVSYLEIYNEVRDL--GKQOQHRLEVKRERPDGVYVKDLSAFV 185

QY 171 VYVSEMALKWITKGEKSRHYGETKMQRSRSHITFIRMILESREKGEPSNCEGSVKVSHL 230

DB 186 VNNADDMDRITMIGNKRSVSGATNMNESSRSHAFITTLERSDMG--LDKEQHRVVGKL 243

QY 231 NLVDLAGSERAQAOTGAAGVRKKEGNCINRSFILGOVVKKLSGQGVGFINRYDSKLTRI 290

DB 244 HMVDLAGSERQDKTGATGDLKEGKIKNLSTALGNVISALVDGK-STHPIRNSKLTRL 302

QY 291 LQNSLGNPKTRIICTITPV--SPDETALQFATAKYMKNTPYVNEVSTDEALLKRYR 348

DB 303 LQDSLGNNAKTMVCANIGPABVYNYDETISTLYANRAKNIKNAKINEDPKD-ALLREFQ 361

QY 349 KEIMDLKKOLEE-----VSLETRAQAM 370

DB 362 KEIEELKKQISEGEGGLDDDEESGESSEGEAGEGKVKKRGKGNPKRLSPKLSPEIMA-AM 420

QY 371 EK--DOLAQLLE-KDLL-----QKVQNEKIENITRMLV 401

DB 421 QKIDEEKKALEEKMDVVEDRNTVHRELQRESELHKAQDDQKILNEKLNIAQKKLI 478

RESULT 11

T06065

hypothetical protein F19H22.150 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 31-Mar-2000

C;Accession: T06065

R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X

submitted to the Protein Sequence Database, March 1999

A;Reference number: 215184

A;Accession: T06065

A;Molecule type: DNA

A;Residues: 1-1121 <BEV>

A;Cross-references: EMBL:AL035679; GSPDB:GN00062; ATSP:F19H22.150

A;Experimental source: cultivar Columbia; BAC clone F19H22

C;Genetics:

A;Gene: ATSP:F19H22.150

A;Map position: 4

A;Introns: 139/2; 170/1; 200/1; 256/3; 284/3; 302/3; 322/3; 349/1; 371/3; 390/3; 425/3;

C;Superfamily: kinesin heavy chain; kinesin motor domain homology

F;99-494/Domain: kinesin motor domain homology <KMOT>

Query Match

Best Local Similarity 33.8%; Score 694; DB 2; Length 1121;

Matches 178; Conservative 87; Mismatches 131; Indels 96; Gaps 14;

QY 1 ABEAGAVAVCVRVRLN-----SRRES-----IGETA---QVYVTKDN 34

DB 94 SERDSISVTVRFRPLRYARSDLAMDQIYCRDYSFHVDAIGVNSLLGGEYQGVDEVAWYPDG 153

QY 35 NVI--YQVDGSKSNFDRVFGHNETTKNVYEIAAIIIDSAIQGYNGTIFAYGQTASGKT 92

DB 154 DTLVRHEYNPLTAVAFDKVFGPQATIDVDVAARPVVKAAMEGVNGTVFAYGVTSSTGKT 213

QY 93 YTM-----MGSEDLHLGVIPRAIHDF 113

DB 214 HTMHVRVLKXKNPYTPPSLVFLQNTCANWDLFFNLVPLVYLLGQESFGIIPLAIKDVF 273

QY 114 QKIKKF-----PDREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVYVADLT 167

DB 274 SLIQDVSGLNGTPGREFLLRVSYLEIYNEVINDLLDPTG--QNLVRRED-SQGTVEGIGK 330

QY 168 BEVVYVTSMAKWIITKGEKSRHYGETKMQRSRSHITFIRMILESREKGEPSNCEGSVKV 227

Db 331 EGVLSFGHALSFIAAGEEHHVGSNNFNLLSSRSHTIFILMVESATGDEYD---GVIF 387

QY 228 SHMLVDLAGSERAQAGVRUKKEGNCINRSFILGQVTKKLSGQGVGFNRYRDSKL 287

Db 388 SOLMLIDLAGE--SSKTETGLRRKEGYSINKSLTLGTVIGLSEK-AATHIPYRDSKL 445

QY 288 TRILONSLGNPKTRIITCTTPV--SFDETLTALQFASAKYMNTPYVNEVSTDEALLK 345

Db 446 TRLOSSLGSHGVSLCTITTPASSSEETHTLKFASRAKSIETIYASRNQIIDEKSLIK 505

QY 346 RYRKEIMDLKQLEEV-----SLETRAQAMEKQD--LAQLEEKDLQKQVNEK 392

Db 506 KYQREISTLKLQDLRGLVGVSHBELSLKQLEEGQVQMQRSLJEEEBEAKALMSR 565

QY 393 IENUTRMLVTSS 404

Db 566 IQKLTKILWST 577

RESULT 12

A54803

N;Alternate names: kinesin-related motor KIF4 - mouse

C;Species: Mus musculus (house mouse)

C;Date: 05-Apr-1995 #sequence_revision 05-Apr-1995 #text_change 02-Feb-2001

C;Accession: A54803; D44259

R;Sekine, Y.; Okada, Y.; Noda, Y.; Kondo, S.; Aizawa, H.; Takemura, R.; Hirokawa, N.

J. Cell Biol. 127, 187-201, 1994

A;Title: A novel microtubule-based motor protein (KIF4) for organelle transports, whose

A;Reference number: A54803; MUID: 95014709; PMID: 7929562

A;Accession: A54803

A;Molecule type: mRNA

A;Residues: 1-1231 <SEQ>

A;Cross-references: GB:D12646; NID:G563772; PIDN:BA002167.1; PID:dh1002657; PID:G563773

R;Aizawa, H.; Sekine, Y.; Takemura, R.; Zhang, Z.; Nangaku, M.; Hirokawa, N.

J. Cell Biol. 119, 1287-1296, 1992

A;Title: Kinesin family in murine central nervous system.

A;Reference number: A44259; MUID: 93077686; PMID: 1447303

A;Accession: D44259

A;Status: nucleic acid sequence not shown; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 91-111,'S',113-240 <AIZ>

A;Experimental source: brain

A;Note: sequence extracted from NCBI backbone (NCBI:P118904)

C;Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology

C;Keywords: ATP; nucleotide binding; P-loop

F;10-343/Domain: kinesin motor domain homology <RMOT>

F;88-95/Region: nucleotide-binding motif A (P-loop)

Query Match 33.8%; Score 693; DB 2; Length 1231;

Best Local Similarity 39.3%; Pred. No. 4.6e-37;

Matches 167; Conservative 69; Mismatches 143; Indels 46; Gaps 10;

QY 6 VAYCVVRPLNSREESLG-ETAQVYKTDNNVYQVDSKSFNDFRVPHGNETTKNVEE 64

Db 10 VRVALRCPLVSKIEKGCQCLSFVPGEPQV--VGNDKSFYDFVFDPESTQEVEFNT 67

QY 65 IAAPIDSAIQGNGTIFAYGQTASGTYTMMGS-----EDH-----LGVIPRAIHDFQKIK 117

Db 68 AVAPLIKGVFKGNATVAYGQTGSGKTYSMGAYTAEQHDSDAISGIVPVIQLLFKEIN 127

QY 118 KPDPRELLRVSYMEIYNETITDLIGT-QKMKPLIREDVNRNVVADLTREVYVTSEM 176

Db 128 KKSDFEFTLKVSYLEIYNEEILDLCSREKATQINIREDPKGIKIVGLTEKTVILVAD 187

QY 177 ALKMTITGKSRHYGTQKMNQSRSHITIFRMILESREKGEPSNCEGSKVSHNLVDLA 236

Db 188 TVSLCEQGNNSRTVASTAMNSQSSRSHAIPTISIEQKK----NDRNSSFRSKLHLVDLA 243

QY 237 GSERAQTAGVRUKKEGNCINRSFILGQVTKKLSGQGVGFNRYRDSKLTRILQNSLG 296

Db 244 GSERQKTKAEGRLREGININRGLLCLGNVISALGDDKGNFVPRYRDSKLTLLQDSLG 303

QY 297 GNPKTRIICITTPV--SFDETLTALQFASAKYMNTPYVNEVSTDEALLKRYRKEIMDL 354

Db 304 GNSHTLMIACVSPADSNLEETLNTLYADRAKRIKPKPIIN-----IDPQAAELNHL 355

QY 355 KQLEEVSL-----ETRAQAMEKDQLAQLLEEKDLQKQVNEKLEN 395

Db 356 KQVQVQIILLQAHGGTLPGDINVEPSENLQSLMEKNQ--SLVEENEKLSRGLSEARQ 413

QY 396 LTRML 400

Db 414 TAQWL 418

RESULT 13

H86350

hypothetical protein F8K7.17 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001

C;Accession: H86350

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Iuros, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.B.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID: 21016719; PMID: 11130712

A;Accession: H86350

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-909 <STO>

A;Cross-references: GB:AE005172; NID:G5263326; PIDN:AAD41428.1; GSPDB:GN00141

C;Genetics:

A;Map position: 1

Query Match 33.7%; Score 692; DB 2; Length 909;

Best Local Similarity 39.3%; Pred. No. 3.5e-37;

Matches 167; Conservative 85; Mismatches 141; Indels 32; Gaps 10;

QY 6 VAYCVVRPLNSREESLG-ETAQVYKTDNNVYQVDSKSFNDFRVPHGNETTKNVEE 65

Db 76 ITVTIFRPLSPREVNGDEIAWYAGDVTIRNEYNPSLCYGFDRVFGPPTTRVVDIA 135

QY 66 AAPIDSAIQGNGTIFAYGQTASGTYTMMGSDEHLGVIPRAIHDFQKIKF-----119

Db 136 AQOVVSGAMSGINGTVFAYGVTSSGKTHMHGQSRSPGIPLAVKDVFSIIQEVIANYS 195

QY 120 ----PDRELLRVSYMEIYNETITDLIGTQKMKPLIREDVNRNVVADLTREVYVTSE 175

Db 196 AIQTPEREFLRVSYLEIYNEIYNEINDLLDPTG--QNLRIRED--SQGTYYVEGKDEVVLSA 252

QY 176 MALKMTITGK---KSRHYGTQKMNQSRSHITIFRMILESREKGEPSNCEGSKVSHNL 232

Db 253 HALSLIASGEVIAEHRHVGNNVNLFSRSHTNFTLTISSPHGKGDGGE-DVLSQLHL 311

QY 233 VDLAGSERAQTAGVRUKKEGNCINRSFILGQVTKKLSGQGVGFNRYRDSKLTRELQ 292

Db 312 IDLAGSE--SKTEITGQRKEGSSINKSLTLGTVISKLTDTK-AAHIFPYRDSKLTRELQ 369

QY 293 NSLGNPKTRIITCTTPVS--FDETLTALQFASAKYMNTPYVNEVSTDEALLKRYRKE 350

Db 370 STLSGHGRVSLCTITTPASSTSEETHTLKFQRCQKVEIKASRNKIMDEKSLIKRYQKE 429

QY 351 IMDLKKOLEEV-----SLETRAQAMEKDQLAQL-----LEEKDLQKQVNEKLENITRM 399

Db 430 ISCLQELTQLRHGNODDLADRLQVGLKSKCVKLQSKRLEDEDEAKAALMGRIQLTKL 489

QY 400 LVTSS 404

Db 490 ILVST 494

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:26:36 ; Search time 3.88047 Seconds
(without alignments)
5421.082 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_405

Perfect score: 2053

Sequence: 1 AEEGAVVCVRVRPLNSREE.....LQKVQNEKIENLTRLMTSS 404

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2053	100.0	2663	1	CENE HUMAN
2	723	35.2	701	1	KF3A_MOUSE
3	716.5	34.9	702	1	KF3A_HUMAN
4	714	34.8	747	1	KF3B_HUMAN
5	713	34.7	747	1	KF3B_MOUSE
6	712.5	34.7	742	1	KI21_STRPU
7	708.5	34.5	785	1	FL10_CHLRE
8	702	34.2	699	1	KI22_STRPU
9	694.5	33.8	1232	1	KF4A_HUMAN
10	693	33.8	1231	1	KF4A_MOUSE
11	691.5	33.7	1029	1	KF17_HUMAN
12	687	33.5	1226	1	KF4A_XENLA
13	683	33.3	1038	1	KF17_MOUSE
14	667	32.5	1225	1	KF4A_CHICK
15	662.5	32.3	1066	1	KL61_DROME
16	656	32.0	672	1	OSM3_CAEEL
17	652	31.8	975	1	KINH_DROME
18	651	31.7	796	1	KF3C_MOUSE
19	649.5	31.6	1749	1	KI3A_MOUSE
20	649	31.6	796	1	KF3C_RAT
21	648.5	31.6	793	1	KF3C_HUMAN
22	645	31.4	1031	1	KINH_STRPU
23	644.5	31.4	1805	1	KI3A_HUMAN
24	643.5	31.3	1826	1	KI3B_HUMAN
25	637	31.0	967	1	KINH_LOLPE
26	634.5	30.9	1032	1	KINH_HUMAN
27	631.5	30.8	957	1	KF5C_HUMAN
28	631.5	30.8	1027	1	KINH_MOUSE
29	630.5	30.7	956	1	KF5C_MOUSE
30	630.5	30.7	963	1	KINH_MOUSE
31	629.5	30.7	963	1	KINH_HUMAN
32	625	30.4	784	1	KL68_DROME
33	622	30.3	815	1	KINH_CAEEL

34	622	30.3	1584	1	UI04_CAEEL
35	620	30.2	554	1	KLP3_SCHPO
36	609.5	29.7	928	1	KINH_NEUCR
37	609.5	29.7	1060	1	EG51_XENLA
38	609	29.7	1067	1	EG52_XENLA
39	606.5	29.5	1057	1	KF11_HUMAN
40	603	29.4	935	1	KINH_SYNRA
41	602.5	29.3	1648	1	KF14_HUMAN
42	601.5	29.3	1111	1	KLP1_YEAST
43	596	29.0	1056	1	KI25_ARATH
44	591.5	28.8	1690	1	KF1A_HUMAN
45	591	28.8	1184	1	BIMC_EWENI

ALIGNMENTS

RESULT 1
CENE_HUMAN
ID CENE_HUMAN STANDARD; PRT; 2663 AA.
AC Q02224;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Centromeric protein B (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J.; Li G.; Schaar B.T.; Szilak I.; Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis.";
RL Nature 359:536-539 (1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A.; Jordan M.A.; Schaar B.T.; Yen T.J.; Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RL microtubule motor.";
RN [3]
RX EMBO J. 14:918-926 (1995).
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T.; Schaar B.T.; Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RL interactions with the kinetochore proteins CENP-F and hBUBR1.";
RN [4]
RX J. Cell Biol. 143:49-63 (1998).
RP FARNESYLATION.
RX MEDLINE=10459117; PubMed=10852915;
RA Ashar H.R.; James L.; Gray K.; Carr D.; Black S.; Armstrong L.;
RT Bishop W.R.; Kirschmeier P.;
RT "Farnesyl transferase inhibitors block the farnesylation of CENP-E
RT and CENP-F and alter the association of CENP-E with the
RL microtubules.";
RN J. Biol. Chem. 275:30451-30457 (2000).
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING
CC CONGRESSION. RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; Z15005; CAA78727.1; -;
DR PIR; S28261; S28261.
DR HSP; P17119; 3KAR.
DR Genew; HGNC:1856; CENPE.
DR GK; Q02224; -;
DR MIM; 117143; -;
DR GO; GO:0005699; C:kinetochore; TAS.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0008350; F:kinetochore motor activity; TAS.
DR GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
DR GO; GO:0007079; P:mitotic chromosome movement; TAS.
DR GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
DR InterPro; IPR001752; Kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere; Lipoprotein; Prenylation.
FT DOMAIN 1 335 KINESIN-MOTOR.
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP BIND 86 93 ATP (BY SIMILARITY).
FT LIPID 2660 2660 S-farnesyl cysteine.
FT SEQUENCE 2663 AA; 312087 MW; CEFC13880C8C8B8 CRC64;

Query Match 100.0%; Score 2053; DB 1; Length 2663;
Best Local Similarity 100.0%; Pred. No. 3.6e-123;
Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AEEGAVAVCVVRPLNSREESLGETAQVYKTDNNVIYQVDSKSFNDFRVFHNETHKN 60
Db 2 AEEGAVAVCVVRPLNSREESLGETAQVYKTDNNVIYQVDSKSFNDFRVFHNETHKN 61
QY 61 VYBEIAAPIIDSAIQGYNGTIFAYGQTASGKTYTMGSEDLGVIPRAIHDIQKIKFP 120
Db 62 VYBEIAAPIIDSAIQGYNGTIFAYGQTASGKTYTMGSEDLGVIPRAIHDIQKIKFP 121
QY 121 DREFLLRVSYMEIYNETITDLCTQKMKPLIREDVNRNVIYADLVEEVVYTMSEALKW 180
Db 122 DREFLLRVSYMEIYNETITDLCTQKMKPLIREDVNRNVIYADLVEEVVYTMSEALKW 181
QY 181 ITKGEKSRHYGETKMNQSRSSHTIFRMILESRKGEPSNCEGSKVSHLNLVDLAGSR 240
Db 182 ITKGEKSRHYGETKMNQSRSSHTIFRMILESRKGEPSNCEGSKVSHLNLVDLAGSR 241
QY 241 AAOCTGAAGVRLKEGCNINRSFLIGQVTKLSDGVGFNYRDSKLTIRLQNSLGNPK 300
Db 242 AAOCTGAAGVRLKEGCNINRSFLIGQVTKLSDGVGFNYRDSKLTIRLQNSLGNPK 301
QY 301 TRITCTTPVSFDETLTALQAFSTAKYMNTPYNEVSTDEALLKRYRKEIMDLKQLEE 360
Db 302 TRITCTTPVSFDETLTALQAFSTAKYMNTPYNEVSTDEALLKRYRKEIMDLKQLEE 361
QY 361 VSLTRAQAMEKDQALQLEEKLLQKQVNEKIENLRLMTVSS 404
Db 362 VSLTRAQAMEKDQALQLEEKLLQKQVNEKIENLRLMTVSS 405

RESULT 2
KF3A MOUSE
ID KF3A MOUSE STANDARD; PRT; 701 AA.
AC P28741;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Kinesin-like protein KIF3A (Microtubule plus end-directed kinesin

DE motor 3A).
GN KIF3A OR KIF3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=93077686; PubMed=1447303;
RA Aizawa H., Sekine Y., Takemura R., Zhang Z., Nangaku M., Hirokawa N.;
RT "Kinesin family in murine central nervous system.";
RL J. Cell Biol. 119:1287-1296(1992).
CC -!- FUNCTION: MICROTUBULE-BASED ANTEROGRADE TRANSLOCATOR FOR
CC MEMBRANOUS ORGANELLES. PLUS END-DIRECTED MICROTUBULE SLIDING
CC ACTIVITY IN VITRO.
CC -!- SUBUNIT: HETERODIMER OF KIF3A AND KIF3B.
CC -!- TISSUE SPECIFICITY: EXPRESSED ALMOST EXCLUSIVELY IN ADULT BRAIN
CC TISSUE (MAINLY IN THE CEREBELLAR GRANULAR LAYER) WITHIN A SINGLE
CC TYPE OF NEURONAL CELL.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC II SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; D12645; BAA02166.1; -;
DR PIR; B44259; B44259.
DR HSP; P17119; 3KAR.
DR MGD; MGI:107689; Kif3a.
DR InterPro; IPR001752; Kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
KW Motor protein; Microtubule; ATP-binding; Coiled coil; Neurone.
KW KINESIN-MOTOR (BY SIMILARITY).
FT DOMAIN 1 350 KINESIN-MOTOR (BY SIMILARITY).
FT DOMAIN 351 598 COILED COIL (BY SIMILARITY).
FT DOMAIN 599 701 GLOBULAR.
FT NP BIND 100 107 ATP (BY SIMILARITY).
FT DOMAIN 442 445 POLY-GLU.
FT DOMAIN 509 512 POLY-ARG.
FT SEQUENCE 701 AA; 80167 MW; 2405872DF2D85A29 CRC64;
Query Match 35.2%; Score 723; DB 1; Length 701;
Best Local Similarity 44.4%; Pred. No. 6.7e-39;
Matches 178; Conservative 66; Mismatches 137; Indels 20; Gaps 10;
QY 6 VAVCVVRPLNSREESL--GETAQVYKTDNNVIYQVDSG----KSFNDFRVFHNETHKN 59
Db 15 KVWVRCRPLNREKSKCYQAVSDVEMRGITVHKTDSSNEPKTFTFDVFPESKQL 74
QY 60 NVYBEIAAPIIDSAIQGYNGTIFAYGQTASGKTYTMGSE---DHLGVIPRAIHDIQKI 116
Db 75 DVYNLTARPIIDSVLEGYNGTIFAYGQTGKTGTMEGVRAVPLRGVIFNPSFAHIFGHI 134
QY 117 KFFP-DREFLLRVSYMEIYNETITDLCTQKMKPLIREDVNRNVIYADLVEEVVYTMSE 175
Db 135 AKAEGRDTRFLVRVSYLEIYNEEVRDLL--GKDTQRLVKERPDVGVYKLSAVVNNAD 193
QY 176 MALKWIITKGEKSRHYGETKMNQSRSSHTIFRMILESRKGEPSNCEGSKVSHLNLVDL 235
Db 194 DMDRIMTLGHKVRSGVATNNWHSRSHAIPTITIECKEKGVDGNMH--VRMGKHLVDL 251
QY 236 AGSERRAQTGAAGVRLKEGCNINRSFLIGQVTKLSDGVGFNYRDSKLTIRLQNSL 295
Db 252 AGSERQANTGATGQRLKEATKINLSLTGLNVI SALVDGK--STHVPYRNSKLTIRLQNSL 310

DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Kinesin-like protein KIF3B (Microtubule plus end-directed kinesin motor 3B) (HH0048).
 GN KIF3B OR KIAA0359.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RA "Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.";
 RT DNA Res. 4:141-150(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R., Jones M., Stavros G., Almeida J.P., Babbage A.K., Bagguley C.L., Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Clee C.M., Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clegg S., Copley V.E., Collier R.E., Connor R.E., Corby N.R., Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Graffam D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E., Hammond S., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Huckle E., Hurley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S., Leheslahti M.H., Levenson M.A., Lloyd C., Lloyd D.M., Lovell J.D., Marsh V.L., Martin S.L., McConnell L.J., McLay K., McMurray A.A., Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T., Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I., Phillips R.B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H., Rice C.M., Ross M.T., Scott C.E., Sehara H.K., Showkneen R., Sims S., Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E., Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A., Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A., Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S., Rogers J.;
 RA "The DNA sequence and comparative analysis of human chromosome 20.";
 RT Nature 414:865-871(2001).
 RN [3]
 RP IDENTIFICATION IN A COMPLEX WITH SMC3 AND KIFAP3B.
 RX MEDLINE=98175913; PubMed=9506951;
 RA Shimizu K., Shiratani H., Honda T., Minami S., Takai Y.;
 RA "Complex formation of SNAP/KAP3, a KIF3A/B ATPase motor-associated protein, with a human chromosome-associated polypeptide.";
 RT J. Biol. Chem. 273:6591-6594(1998).
 CC -!- FUNCTION: Involved in tethering the chromosomes to the spindle pole and in chromosome movement. Microtubule-based anterograde translocator for membranous organelles. Plus end-directed microtubule sliding activity in vitro (By similarity).
 CC -!- SUBUNIT: Heterodimer of KIF3A and KIF3B (By similarity). Interacts with the SMC3 subunit of the cohesin complex.
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN II SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AB002357; BAA20815.1; -;
 DR EMBL; AL121897; CAC16425.1; -;
 DR HSSP; P17119; 3KAR.
 DR Genew; HGNC:6320; KIF3B.
 DR MIM; 603754; -;
 DR GO; GO:0005873; C:plus-end kinesin complex; TAS.
 DR GO; GO:0003777; F:Microtubule motor activity; TAS.
 DR GO; GO:0008574; F:plus-end-directed kinesin ATPase activity; TAS.
 DR GO; GO:0008089; P:anterograde axon cargo transport; TAS.
 DR GO; GO:0007368; P:determination of left/right asymmetry; TAS.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; Coiled coil; Neurone.
 FT DOMAIN 1 345 KINESIN-MOTOR (BY SIMILARITY).
 FT DOMAIN 346 579 COILED COIL (BY SIMILARITY).
 FT DOMAIN 580 747 GLOBULAR.
 FT NP BIND 96 103 ATP (POTENTIAL).
 FT DOMAIN 386 393 POLY-GLY.
 FT DOMAIN 394 406 POLY-GLU.
 FT DOMAIN 723 730 POLY-SER.
 FT SEQUENCE 747 AA; 85125 MW; 97FA4573AFA87023 CRC64;
 SQ
 Query Match 34.8%; Score 714; DB 1; Length 747;
 Best Local Similarity 41.4%; Pred. No. 2.7e-38;
 Matches 177; Conservative 69; Mismatches 144; Indels 38; Gaps 10;
 QY 5 AVAVCVVRPLNSRESLG-----ETAVYWKTDNNVIYQVDSKSFNDRVPHGN 55
 Db 9 SVRVVRCRPMNGEKAASYDKVVDVVLGVQSVNPKGTAHEM--PKTFTFDDAVTDWN 66
 QY 56 ETTKNVYEETAAPIIDSAIOGYNGTTFAYGQTASGKTYTM---MGSEDLHLGVTPRAIHD 112
 Db 67 AKOFELYDETRFLVDSVLQGFNGTTFAYGQTGKTYTMEGIRGDPEKRGVTPNSFDHI 126
 QY 113 FQKIKPEPDRFLIRVSYMEIYNETITDLLCGTQMKPLIREDVNRNVVADLTVEVY 172
 Db 127 FTHISRQOQYLRASYLEIYBEIRDLLSKDQ--TKRLBELKRPDTGVYVKDLSFFVTK 185
 QY 173 TSEMAKWIYKGEKSRHYGTGKQNRSSRSHITFRMILESEKGEPCSVKSHNL 232
 Db 186 SVKEIEHVMVGNQNSVGATNMNEHSRSHAFVITIECEVG--LDGENHVRGKLN 243
 QY 233 VDLAGSRAAQTGAAGVRLKEGCMNRSFLTGQVKKLSDGVGGGFTNYRDSKLRLQ 292
 Db 244 VDLAGSERQKTGAQGERLKEATKINLSLGNVISALVDGK--STHPIYRDSKLRLQ 302
 QY 293 NSLGGNPKRICTITPVSF--DETLALQFASAKYMKNTPPYNEVSTDEALLKRYKE 350
 Db 303 DSLGNAKTVMVANVPASYNVEETLTLRYANRAKIKKPKRVNEDPKD--ALLREFOEE 361
 QY 351 IMDLKKQLEEVSLSTRAQ-----AMEKDLQALLEE---KDLLQKVQNEKI 393
 Db 362 IARLKAQLKRSIGRRKRKRREGGGGGGGGEGEGEGEGEGEGEGEGEGEGEGEGEG 421
 QY 394 ENLTRLV 401
 Db 422 EIEKRAIV 429
 RESULT 5
 KF3B_MOUSE
 ID KF3B_MOUSE STANDARD; PRT; 747 AA.
 AC Q61771;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Kinesin-like protein KIF3B (Microtubule plus end-directed kinesin motor 3B).

GN KIF3B.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR; TISSUE=Brain;
 RX MEDLINE=96032268; PubMed=7559760;
 RA Yamazaki H., Nakata T., Okada Y., Hirokawa N.;
 RT "KIF3A/B: a heterodimeric kinesin superfamily protein that works as a
 RT microtubule plus end-directed motor for membrane organelle
 RT transport."; J. Cell Biol. 130:1387-1399(1995).
 RL J. Cell Biol. 130:1387-1399(1995).
 CC -!- FUNCTION: Involved in tethering the chromosomes to the spindle
 CC pole and in chromosome movement. Microtubule-based anterograde
 CC translocator for membranous organelles. Plus end-directed
 CC microtubule sliding activity in vitro (By similarity).
 CC -!- SUBUNIT: Interacts with the SMC3 subunit of the cohesin
 CC complex (By similarity). Heterodimer of KIF3A and KIF3B.
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
 CC II SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; D26077; BAA05070.1; -
 DR PIR; A57107; A57107.
 DR HSSP; P17119; 3KAR.
 DR MGD; MGI:107688; Kif3b.
 DR InterPro: IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINSEINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; Coiled coil; Neurone.
 FT DOMAIN 1 345 KINESIN-MOTOR (BY SIMILARITY).
 FT COILED COIL (BY SIMILARITY).
 FT GLOBULAR.
 FT NP BIND 96 103 ATP (POTENTIAL).
 FT DOMAIN 386 393 POLY-GLY.
 FT DOMAIN 394 405 POLY-GLU.
 FT DOMAIN 723 730 POLY-SER.
 SQ SEQUENCE 747 AA; 85288 MW; FA369A4190EC8B47 CRC64;
 Query Match 34.7%; Score 713; DB 1; Length 747;
 Best Local Similarity 41.4%; Pred. No. 3.2e-38;
 Matches 177; Conservative 69; Mismatches 144; Indels 38; Gaps 10;
 QY 5 AVAVCVVRPLNSREESLG-----ETAQVYWKTDNNVIYQVDSKGFNFDRVPHGN 55
 DB 9 SVRVVVRCPMGKEKAASYKVDVVKLGQVSKNPKGTSHM--PKTFTFVAVDWN 66
 QY 56 ETTKNVVEEIAAPIIDSAIOGYNGTIFAYGQTASGKTYT---MGSEDHGLVIPRAIHI 112
 DB 67 AKOFELYDETFRLPLVDVSLQGFNGTIFAYGQTGTGKTYTMEGVGRDPEKRGVIPSFDHI 126
 QY 113 FQIKKPPDFRFLRVSMYEINETITDLCCGTQMKPLIIRVDNVRNVVADLTVEVY 172
 DB 127 FTHLSRQOQVLRASYLIYQBEIRDLSSKQ--TKRLKRPOTGVVVKDLSSEVTK 185
 QY 173 TSEWALKWITKESRHYGTHKQNRSSRHTIFRMILSRKSGPNCVGVKVSHLN 232
 DB 186 SVKEIEHVMVNGVQNRSGVATNMNHSRSHAFVITIECEVG--LDGENHVRVGLKL 243
 QY 233 VDLAGSERAQTCAGVRLKEGNCNINSLFILGVIKLSDGVGGVFINTYRDSKLTRELQ 292

Db 244 VDLAGSERAQTCAGQGERLKEATKINLSLSALGNVISALVDGK-STHPIYRDSKLTRELQ 302
 QY 293 NSLGGNPKTRIICTIPVVF--DETALTALQFASAKYMKNTPYVNEVSTDEALLKRYRKE 350
 DB 303 DSLGGNAKTMVAVNGVPASVYVEETLTLLRYANRAKNKKNRVNEDPKD-ALLREFQEE 361
 QY 351 IMDLKKOLEEVSLETRAQ-----AMEKDQLAQLLEE-----KOLLOKVONEKI 393
 DB 362 IARLKAQLEKRSIGRRKRREKRREGSGGGGEGEEGEGEDGDDKDDYVREQQEK 421
 QY 394 ENLTRLIV 401
 DB 422 EIEKRAIV 429
 RESULT 6
 ID K121_STRPU STANDARD; PRT; 742 AA.
 AC P46871;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Kinesin-II 95 kDa subunit (KRP-85/95 95 kDa subunit).
 GN KRP95.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
 OC Strongylocentrotus.
 OX NCBI_TaxID=7668;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Egg;
 RX MEDLINE=94050179; PubMed=8232586;
 RA Cole D.G., Chinn S.W., Wedaman K.P., Hall K., Vuong T.,
 RA Scholey J.M.;
 RT "Novel heterotrimeric kinesin-related protein purified from sea
 RT urchin eggs";
 RL Nature 366:268-270(1993).
 CC -!- SUBUNIT: Heterotrimer of a 115 kDa subunit (KAP115) and two
 CC kinesin-like subunits of 95 kDa (KRP95) and 85 kDa (KRP85).
 CC -!- PTM: The N-terminus is blocked.
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
 CC II SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U00996; AAA87393.1; -
 DR HSSP; P17119; 3KAR.
 DR InterPro: IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINSEINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; Coiled coil.
 FT DOMAIN 1 337 KINESIN-MOTOR (BY SIMILARITY).
 FT COILED COIL (BY SIMILARITY).
 FT GLOBULAR (BY SIMILARITY).
 FT NP BIND 95 102 ATP (POTENTIAL).
 SQ SEQUENCE 742 AA; 84202 MW; 47C40A367BAA77B5 CRC64;
 Query Match 34.7%; Score 712.5; DB 1; Length 742;
 Best Local Similarity 41.3%; Pred. No. 3.4e-38;
 Matches 175; Conservative 65; Mismatches 137; Indels 47; Gaps 10;
 QY 6 VAVCVVRPLNSREESLGETAQVYWKTDNNVIYQV-----DGSKSFNDRVPHGN 58

9 VKVVRCPMNSKETSQCHKRIVEMDNKRGLVEVTPKPPGPPNPKSFTFDVYDWSKQ 68
 59 KNYVEIAPIDSAIOGNGTIFAYGOTASCTVTMMG--SEDLGVIPRAIHDFQK 115
 69 IDLYDETFRSLVESVLOQNGTIFAYGOTGKTFTMEGVRNPELRGVIPNPFHIFH 128
 116 IKKFPDRFLLVSVMEIYNETITDLCCTQKMKPLIREDVNRNYYVADLTFEYVYTS 175
 129 IARTNQQLVRAVSLVLEIYQEBIRDLAKDQK-KRLDLKERPDGTGVYVVDLSFVTKSVK 187
 176 MALKWITKGEKSRHYGETKQNRSSRSHTIFRMILESREKGFNSCEGSKVYSHNLVDL 235
 188 EIEHVMTVGNNSRVSGTNNHESRSHAIFITIECSLGVGD--ENHIRVGKULVDL 245
 236 AGSERAQACAGVRLKEGNCNRSFILGOVTKLSDGQVGFINVRDSKLTIRIQLNSL 295
 246 AGSERAQKATGATGDRLEKATKINLSALGNVISALVDGK-SSHIPYRDSKLTIRIQLNSL 304
 296 GGNPKTRIICTITPVV--FDETLTALQFASCTAKYKNTPYVNEVSTDEALLKRYRKEIMD 353
 305 GGNKATVWVNMGPASVNFDETTITLRYANRAKNIKNPKINEDPKD-ALLREFQBEISR 363
 354 LKQQL-----EVSLTETRAQAMEKDQLAQLEE--KDLLOKV 388
 364 LKQALDKGSPDGRKKGKPKGQGGDDDIREDTEEBGEMDE-----EEMTKESQOQKL 418
 389 QNEK 392
 419 EEEK 422

RESULT 7
 FL10_CHRE
 ID FL10_CHRE STANDARD; PRT; 786 AA.
 AC P46869;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Kinesin-like protein FLA10 (KHP1 protein).
 GN FLA10.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=137;
 RA MEDLINE=94299638; PubMed=8027176;
 RA Walther Z., Vashishtha M., Hall J.L.;
 RT "The Chlamydomonas FLA10 gene encodes a novel kinesin-homologous protein";
 RL J. Cell Biol. 126:175-188(1994).
 CC -!- FUNCTION: Probably involved in flagellar assembly and maintenance.
 CC May play a role in flagellar synthesis.
 CC -!- TISSUE SPECIFICITY: Flagellar axoneme.
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
 CC II SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L33697; AAA21738.1; -;
 DR PIR; A53939; A53939.
 DR HSP; P17119; 3KAR.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.

DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE; PS00667; KINESIN_MOTOR_DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; Coiled coil.
 FT DOMAIN 1 358 KINESIN_MOTOR (BY SIMILARITY).
 FT DOMAIN 367 687 COILED COIL (POTENTIAL).
 FT DOMAIN 688 786 GLOBULAR (POTENTIAL).
 FT NP_BIND 97 104 ATP (POTENTIAL).
 FT DOMAIN 388 391 POLY-GLY.
 FT DOMAIN 705 714 POLY-GLY.
 FT DOMAIN 756 759 POLY-ASP.
 SQ SEQUENCE 786 AA; 86671 MW; F90969203EB79F1B CRC64;
 Query Match 34.5%; Score 708.5; DB 1; Length 786;
 Best Local Similarity 44.4%; Pred. No. 6.6e-38;
 Matches 167; Conservative 66; Mismatches 118; Indels 25; Gaps 9;
 QY 5 AVAVCVVRPLNRSREESLGETAQVWKTNNVIY-----QVDGS---KSFNFRVFGHNET 57
 10 SVKVVVRCPLNGKEKADGRSRIVDMVDAGVKVKNPKADASEPPKAFTFDQVYDWNQC 69
 58 TKNVYEEIAPIDSAIOGNGTIFAYGOTASCTVTMMGSEDH---LGVIPRAIHDFQ 114
 70 QRDVFDITARPLDSCIEGTNGIIFAYGOTGKSHTEGKDEPPELRGLIPNTRYVFE 129
 115 KI-KKFPDRFLLVSVMEIYNETITDLCCTQKMKPLIREDVNRNYYVADLTFEYVYT 173
 130 IIRDSGTKEFLVRSYLYEINEVRDLL-GKDHKKMELKESPDGRGVYVVDLSQFVCKN 188
 174 SEMALKWITKGEKSRHYGETKQNRSSRSHTIFRMILESREK-----GEPSNCEGS 224
 189 YEEMNKVLLAGKDNQVQATLMNQDSSRSHTIFITIECTEKLESAAAPGAKKDDSNH 248
 225 VKVSHNLVDLAGSERAQACAGVRLKEGNCNRSFILGOVTKLSDGQVGFINVRD 284
 249 VRVQKLNVDLAGSERQDKTGATGDRLEKIGIKINLSLTALGNVISALVDGK-SGHI 307
 285 SKLTRIILQSLGNPKTRIICTITPV--SFDETLTALQFASCTAKYKNTPYVNEVSTDEA 342
 308 SKLTRIILQSLGNPKTRIICTITPV--SFDETLTALQFASCTAKYKNTPYVNEVSTDEA 342
 343 LLKRYRKEIMDLKQQL 358
 367 MLRFQFEIEIKLKEQL 382

RESULT 8
 KI22_STRPU
 ID KI22_STRPU STANDARD; PRT; 699 AA.
 AC P46872;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Kinesin-II 85 kDa subunit (KRP-85/95 85 kDa subunit).
 GN KRP85.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
 OC Strongylocentrotus.
 OX NCBI_TaxID=7668;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Egg;
 RX MEDLINE=94050179; PubMed=8232586;
 RA Cole D.G., Chinn S.W., Wedaman K.P., Hall K., Vuong T.,
 RA Scholey J.M.;
 RT "Novel heterotrimeric kinesin-related protein purified from sea
 RT urchin eggs";
 RL Nature 366:268-270(1993).
 CC -!- SUBUNIT: Heterotrimer of a 115 kDa subunit (KAP115) and two
 CC kinesin-like subunits of 95 kDa (KRP95) and 85 kDa (KRP85).
 CC -!- PTM: The N-terminus is blocked.
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
 CC II SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; LI6993; AAA16098.1; -;
 CC PIR; S38982; S38982.
 CC HSSP; P17119; 3KAR.
 CC InterPro; IPR001752; kinesin_motor.
 CC Pfam; PF00225; kinesin; 1.
 CC PRINTS; PR00380; KINESINHEAVY.
 CC SMART; SM00129; KISC; 1.
 CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 CC PROSITE; PS50067; KINESIN MOTOR DOMAIN2; 1.
 CC Motor protein; Microtubule; ATP-binding; Coiled coil.
 CC KINESIN-MOTOR (BY SIMILARITY).
 CC COILED COIL (BY SIMILARITY).
 CC GLOBULAR (BY SIMILARITY).
 CC NP BIND 97 104 ATP (POTENTIAL).
 CC SEQUENCE 699 AA; 783666111CB08190 CRC64;
 CC
 CC Query Match 34.2%; Score 702; DB 1; Length 699;
 CC Best Local Similarity 39.7%; Pred. No. 1.5e-37;
 CC Matches 190; Conservative 65; Mismatches 131; Indels 92; Gaps 16;
 CC
 CC 6 VAVCVVRPLNSRESLG-----ETAAQVYWKTDNNVIYVDGSKSFNFRVPHG 54
 CC 11 VRVVRCPPLNSKGTGGQFSKVRKDEMRGTGVQV---TNPNA-PSGEPPKSTFTDVEAP 66
 CC
 CC 55 NETTKNVEEIAAPITDSIAIOQYNGTIFAYGTASGKTYTWG--SEDLH-CVIPRAIHD 111
 CC 67 GAKQTDVYNQTPRIVDAIIQYNGTIFAYGTGKTYTWG--SEDLH-CVIPRAIHD 111
 CC
 CC 112 IFQIKKPPDR-EFLRVSMEIYNETITDLCGTQKMKPLIIRDVNRVNVVADLTREV 170
 CC 127 IFGHIAREQENVRFLVRVSYLEINEEVKDLL-GRDQQRLEVRKPDVGVYVVDLSAFV 185
 CC 171 VYTESMALKWLTKEKSHYGETKMNQSSSHITFRMILSRKGEPSNCEGSKVYVSHL 230
 CC 186 VNNADDMRIMILGNKNSVATNNSSSSKSHAITFTITLERSDMG--LDKEQHVVRVGL 243
 CC 231 NLVDLAGSRAAQTGAAGVRLKEGNCINRSFILGQVTKLSDGQVGGFINVRDSKLTRI 290
 CC 244 HMVDLAGSERQTKTGATGQRLKEATKINLSLTGNVLISSLVGK-STHIIYRNSKLTRL 302
 CC 291 LQNSLGNPKTRITCTTPV--SFDETALQPASTAKYKNTPYVNEVSTDEALLKRYR 348
 CC 303 LQDSLGNNAKTVMCANIPAEYNYDETISTLYRANRAKNIKAKINEDPKD-ALLREFQ 361
 CC 349 KETMDLKKQLEE-----VSLETRAQAM 370
 CC 362 KETEELKKQISEGEGDDEESGSESGDEEAGEGVKKRKGKPKRLKSPEIMA-AM 420
 CC 371 EK--DQLAQLEE-KDLL-----QKVNKEIKENTLRLV 401
 CC 421 QKKIDEKKALEEKQNVEDRNTVHRELQRESELHKAQDDQKILNEKLNIAQKKLI 478
 CC
 CC RESULT 9
 CC KF4A HUMAN STANDARD; PRT; 1232 AA.
 CC AC Q95239; Q9NV6; Q9NV24; Q9UM43;
 CC DT 15-JUL-1999 (Rel. 38, Created)
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
 CC DE 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DT Chromosome-associated kinesin KIF4A (Chromokinesin).
 CC GN KIF4A OR KIF4.
 CC OS Homo sapiens (human).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymphocytes;
 RA Willard L.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RX MEDLINE=20435301; PubMed=10978527;
 RA Oh S.J., Hahn H., Torrey T.A., Shin H., Choi W., Lee Y.M.,
 RA Morse H.C. III, Kim W.;
 RT "Identification of the human homologue of mouse KIF4, a kinesin
 RT superfamily motor protein.";
 RL Biochim. Biophys. Acta 1493:219-224 (2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Retinoblastoma;
 RA Rentach A., Neumann T., Rommerskirch W.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 128-1232 FROM N.A.
 RC TISSUE=Retinoblastoma;
 RX MEDLINE=97311419; PubMed=9168136;
 RA Yan R.-T., Wang S.-Z.;
 RL "Increased chromokinesin immunoreactivity in retinoblastoma cells.";
 RT Gene 189:263-267 (1997).
 CC -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR
 CC SPINDLE STABILIZATION (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic
 CC chromosomes (BY similarity).
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEMATOPOIETIC TISSUES,
 CC FETAL LIVER, SPLEEN, THYMUS AND ADULT THYMUS AND BONE MARROW.
 CC LOWER LEVELS ARE FOUND IN HEART, TESTIS, KIDNEY, COLON AND LUNG.
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
 CC Chromokinesin subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; AF179308; AAD51855.1; -;
 CC EMBL; AF071592; AAD05492.2; -;
 CC EMBL; AJ271784; CAB75427.1; -;
 CC EMBL; AF277375; AAF86334.1; -;
 CC HSSP; P17119; 3KAR.
 CC Genew; HGNC:13339; KIF4A.
 CC GO; GO:0005737; C:cytoplasm; TAS.
 CC GO; GO:0005876; C:spindle microtubule; TAS.
 CC GO; GO:0003777; F:microtubule motor activity; TAS.
 CC GO; GO:0008089; P:anterograde axon cargo transport; TAS.
 CC GO; GO:0006996; P:organelle organization and biogenesis; TAS.
 CC InterPro; IPR001752; kinesin_motor.
 CC Pfam; PF00225; kinesin; 1.
 CC PRINTS; PR00380; KINESINHEAVY.
 CC SMART; SM00129; KISC; 1.
 CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 CC PROSITE; PS50067; KINESIN MOTOR DOMAIN2; 1.
 CC Motor protein; Microtubule; ATP-binding; DNA-binding;
 CC Nuclear protein; Coiled coil.
 CC KINESIN-MOTOR.
 CC COILED COIL (BY SIMILARITY).
 CC GLOBULAR.
 CC ATP (POTENTIAL).
 CC R -> G (IN REF. 2).
 CC S -> T (IN REF. 4).
 CC V -> A (IN REF. 2).
 CC L -> W (IN REF. 2).
 CC L -> H (IN REF. 4).
 CC

FT CONFLICT 564 564 L -> P (IN REF. 2).
FT CONFLICT 600 600 K -> E (IN REF. 3).
FT CONFLICT 668 668 R -> K (IN REF. 3 AND 4).
FT CONFLICT 928 928 Q -> P (IN REF. 1).
FT CONFLICT 958 958 Q -> R (IN REF. 3).
FT CONFLICT 960 960 L -> Q (IN REF. 1).
FT CONFLICT 996 997 LL -> S (IN REF. 4).
FT CONFLICT 1003 1014 QKHLPKDTLSP -> RTLPRIPFYIQ (IN REF. 4).
FT CONFLICT 1022 1022 P -> Q (IN REF. 2).
FT CONFLICT 1077 1077 K -> N (IN REF. 2).
FT CONFLICT 1138 1138 G -> S (IN REF. 2).
SQ SEQUENCE 1232 AA; 139908 MW; FF74052A17A8E87 CRC64;

Query Match 33.8%; Score 694.5; DB 1; Length 1232;
Best Local Similarity 39.8%; Pred. No. 9.3e-37; Indels 45; Gaps 8;
Matches 168; Conservative 66; Mismatches 145;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDRVFHGNKTTKNVYEE 64
DB 10 VRVALRCRPLVSKKEIKGECQCLSFVPGEPOVV-VGNDKSFYDFVDPSTQEVEVT 67
QY 65 IAAPILDSAIQYNGTIFAYGQTASGKTYTMWG-----SEDLGVIPRAIHDIQKIK 117
DB 68 AVAPLIKGVKYNATVLAAYGQTGSGKTYSMGGAYTAQENBPTGVIPRVIQLLFKID 127
QY 118 KPWDRFLLRVSYMEIYNETITDLCTGQKMKPLIREDVNRNYYVADLTTEEYVYTSMA 177
DB 128 KKSDFEFTLVSKYLYEINNEELTLRYADRARKINKKPIV-----IDPQAEELNHL 187
QY 178 LKWTITKEKSRHYGETKMNORSRSHITFRMILESREKGPSCNCEGVKVSHLNLDVLAG 237
DB 188 VSCLEQGNNSRTVASTAMNSQSSRSHAFITISIEQRKSD-----KNSFSRSLKHLVDLA 243
QY 238 SERAAQTGAAGVRLKEGCNINRSLFILGOVTKKLSDCGVGFNYRDSKLTILONSLGG 297
DB 244 SERQKTKAEGDRLEKGININRGLLCLGNVISALGDDKGGFVPRDSKLTILONSLGG 303
QY 298 NPKTRITCTIPV--SPDETALQAFSTAKYMKNTPYVNEVSTDEALLKRYKREIMDLK 355
DB 304 NSHTLMTACVSPADSNLEETLNTLRYADRARKINKKPIV-----IDPQAEELNHL 355
QY 356 KQLEBEVL-----ETRAQMEKQQLAQLLEKDLQKQVNEKLENL 396
DB 356 QVQQLQVLLLAHQHGTLPGSITVRPSENLSIMEKQ--SLVEENKLSRGLSEAAQGT 413
QY 397 TRML 400
DB 414 AQML 417

RESULT 10
KF4A MOUSE STANDARD; PRT: 1231 AA.
ID AC P33174;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chromosome-associated kinesin KIF4A (Chromokinesin).
GN KIF4A OR KIF4 OR KNS4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
SPECIFICITY.
RC STRAIN=ICR; TISSUE=Brain;
RX MEDLINE=95014709; PubMed=7929562;
RA Sekine Y., Okada Y., Noda Y., Kondo S., Aizawa H., Takemura R.,
Hirokawa N.;
RT "A novel microtubule-based motor protein (KIF4) for organelle
transports, whose expression is regulated developmentally";
RL J. Cell Biol. 127:187-201(1994).

RP SEQUENCE OF 91-240 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93077686; PubMed=1447303;
RA Aizawa H., Sekine Y., Takemura R., Zhang Z., Nangaku M.,
Hirokawa N.;
RT "Kinesin family in murine central nervous system";
RL J. Cell Biol. 119:1287-1296(1992).
CC -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR
SPINDLE STABILIZATION.
CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic
chromosomes.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN PYRAMIDAL CELLS IN JUVENILE
HIPPOCAMPUS, GRANULAR CELLS IN JUVENILE CEREBELLAR CORTEX AND IN
ADULT SPLEEN.
CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
Chromokinesin subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC
CC EMBL; D12646; BAA02167.1; --
DR PIR; A54803; A54803.
DR HSSP; P17113; 3KAR.
DR MGD; MGI:108389; Kif4.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; Kisc; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN; 1.
DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
DR Motor protein; Microtubule; ATP-binding; DNA-binding;
KW Nuclear protein; Coiled coil.
KW DOMAIN 1 350 KINESIN-MOTOR.
FT DOMAIN 351 1000 COILED COIL (BY SIMILARITY).
FT DOMAIN 1001 1231 GLOBULAR.
FT NP BIND 88 95 ATP (POTENTIAL).
FT FT CONFLICT 112 112 I -> S (IN REF. 2).
SQ SEQUENCE 1231 AA; 139551 MW; F34F2C2D21158FE4 CRC64;

Query Match 33.8%; Score 693; DB 1; Length 1231;
Best Local Similarity 39.3%; Pred. No. 1.2e-36;
Matches 167; Conservative 69; Mismatches 143; Indels 46; Gaps 10;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDRVFHGNKTTKNVYEE 64
DB 10 VRVALRCRPLVSKKEIKGECQCLSFVPGEPOVV-VGNDKSFYDFVDPSTQEVEVT 67
QY 65 IAAPILDSAIQYNGTIFAYGQTASGKTYTMWG-----EDH---LGVIPRAIHDIQKIK 117
DB 68 AVAPLIKGVKYNATVLAAYGQTGSGKTYSMGGAYTAQENBPTGVIPRVIQLLFKIN 127
QY 118 KPWDRFLLRVSYMEIYNETITDLCTGQKMKPLIREDVNRNYYVADLTTEEYVYTSMA 176
DB 128 KKSDFEFTLVSKYLYEINNEELTLRYADRARKINKKPIV-----IDPQAEELNHL 187
QY 177 ALKWTITKEKSRHYGETKMNORSRSHITFRMILESREKGPSCNCEGVKVSHLNLDVLA 236
DB 188 TVSCLEQGNNSRTVASTAMNSQSSRSHAFITISIEQRKSD-----KNSFSRSLKHLVDLA 243
QY 237 GSERAAQTGAAGVRLKEGCNINRSLFILGOVTKKLSDCGVGFNYRDSKLTILONSLGG 296
DB 244 GSERQKTKAEGDRLEKGININRGLLCLGNVISALGDDKGGFVPRDSKLTILONSLGG 303
QY 297 GNPKTRITCTIPV--SPDETALQAFSTAKYMKNTPYVNEVSTDEALLKRYKREIMDL 354
DB 304 GNSHTLMTACVSPADSNLEETLNTLRYADRARKINKKPIV-----IDPQAEELNHL 355

QY 355 KQLEEVSL-----ETRAQWMEKQQAQLLEEKLLQKQVQNEKIEN 395
 Db 356 KQVQQLQILLQAHGGLTFCIDINVEPSENLSQILMEKNQ--SLVEENKLSRGSLSEAGQ 413
 QY 396 LTRML 400
 Db 414 TAQWL 418

RESULT 11
 KF17 HUMAN
 ID KF17 HUMAN STANDARD; PRT; 1029 AA.
 AC Q9P2E2; Q9S077; Q9N411;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Kinesin-like protein KIF17 (KIF3-related motor protein).
 GN KIF17 OR KIF3X OR KIAA1405.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kimberley A., White S.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 RV [2]
 RP SEQUENCE OF 78-1029 FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RA Ohara O., Nagase T., Kikuno R.;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 RV [3]
 RP SEQUENCE OF 239-1029 FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RX MEDLINE=20181126; PubMed=10718198;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirokawa M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XVI.
 RT The complete sequences of 150 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 7:65-73(2000).
 RV [4]
 RP SEQUENCE OF 618-1029 FROM N.A. (ISOFORM 2).
 RC TISSUE=Pancreas;
 RX MEDLINE=22389257; PubMed=12477932;
 RA Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Raha S.S., Loquellano N.A., Toshiyuki S., Carninci P., Prange C.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Abramson R.D., Mullahy S.J.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting R., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RV [5]
 RP SEQUENCE OF 788-1029 FROM N.A.
 RA Nagata K., Puls A., Futter C., Aspenstrom P., Hall A.;
 RT "The MAP kinase kinase kinases MK2 and MK3 are targets for RAC/Cdc42
 RT and components of microtubule motor complexes.";
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Transports vesicles containing N-methyl-D-aspartate
 CC (NMDA) receptor 2B along microtubules (By similarity).
 CC -1- SUBUNIT: Interacts with LIN-10 PDZ domain (By similarity).

CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9P2E2-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9P2E2-2; Sequence=VSP_008218;
 CC Note=Sequence incomplete. No experimental confirmation
 CC available;
 CC -1- SIMILARITY: Belongs to the kinesin-like protein family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).

CC EMBL; AL663074; -; NOT ANNOTATED CDS.
 CC EMBL; AL391357; -; NOT ANNOTATED CDS.
 CC EMBL; AB037826; BAA92643.2; ALT_INIT.
 CC EMBL; BC036871; AAH36871.1; -;
 CC EMBL; AF009624; AAD01428.1; -;
 CC HSSP; P33176; 1BG2.
 CC Genew; HGNC:19167; KIF17.
 CC MIM; 605037; -;
 CC InterPro; IPR001752; kinesin_motor.
 CC Pfam; PF00225; kinesin; 1.
 CC SMART; SM00129; KISC; 1.
 CC PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 CC PROSITE; PS0067; KINESIN_MOTOR_DOMAIN2; 1.
 CC Motor protein; Microtubule; ATP-binding; Coiled coil; Transport;
 CC Protein transport; Alternative splicing.
 CC DOMAIN 1 265 KINESIN-MOTOR.
 CC COILED COIL (POTENTIAL).
 CC COILED COIL (POTENTIAL).
 CC NP_BIND 91 98 ATP (POTENTIAL).
 CC VARSPPLIC 619 673 VEKLARLSSTVARTDAPQADVPKVPQVQVPAFTDLLEPSDA
 CC RPEAAADFFPRP -> INRACFCVTGVPELVICLSGLQ
 CC CAEWPRGLSHLGAVGGGGLPTSTWVSLHTQ (in
 CC isoform 2).
 CC /FTID=VSP_008218.
 CC SDEDS -> IGRN (IN REF. 5).
 CC MISSING (IN REF. 4).
 CC D -> Y (IN REF. 5).
 CC R -> W (IN REF. 2 AND 3).
 CC THNSPGLSCPLNNSAI -> HTSQLATRPQLTQOQLC
 CC H (IN REF. 5)
 CC SEQUENCE 1029 AA; EC52E8B84AD18656 CRC64;
 CC
 CC Query Match 33.7%; Score 691.5; DB 1; Length 1029;
 CC Best Local Similarity 41.0%; Pred. No. 1.1e-36;
 CC Matches 171; Conservative 62; Mismatches 139; Indels 45; Gaps 11;
 QY 5 AVAVCVVRPLNSRELSIGETAQVYKTD-----NNVIYQVDSKSFNDRVFHNE 56
 Db 5 AVKVVRCRPMNQREEL--RCQPVTVDCARACCCQNFCAADEPPKQTFDGAHVHD 62
 QY 57 TTKNVYEEIAAPLIDSAIQYNGTIFAYGQTASGKTYTMMGSED--HLGVIPRAIHDF 113
 Db 63 VTEQIYNEIAYPLVEGVTEGYNGTIFAYGQTSGKSTMQGLDPPPSQRGIIIPRAFHEVF 122
 QY 114 QTKIKPDRFLRVSYMEIYNETITDLLCGTQMKPLIIREVNRNRYVADLTFEEVYVT 173
 Db 123 ESVQCAENTKFLVRASYLEIYNEDVRDLLGADTKQK-LLEKHEPKGVYKGLSMHTVHS 181
 QY 174 SENALKWIKGSKRHYGETKKNQSRSHITFIRMILES---EKGPSNCEGSVKVSHL 230
 Db 182 VAOCEHIMETGWNRSVGYTLMNKDSSRSHTSIPTISIEMSAVDERG-----KDLHAGKL 236
 QY 231 NLVDLAGSRAAQTAGAVRLKGCNINRSIFLIGQVIKKLSQGVGGFNYNEDSKLTRI 290
 Db 237 NLVDLAGSRSQKGTGATGERLKEATKINLSLALGNVISALVDGRC-KHVPYRDSKLTRL 295

Fri Aug 6 10:49:10 2004

Query Match 33.5%; Score 687; DB 1; Length 1226;
 Best Local Similarity 38.8%; Pred. No. 2.8e-36;
 Matches 178; Conservative 70; Mismatches 145; Indels 66; Gaps 14;

2 BEG-AVAVCVKRLPNSREESLG-ETAQVYWKTDNNVYQVDSKSFNDRVFGHNETTK 59
 4 DEGIPYVALRCPLVPKNNESGKMCCLTFVPEQCVI--VGTESKSTFYDYVDFPSAEQE 61
 60 NYVEIAPIIDSAIOGNYGTTFATGOTASGKTYTMMGSEDH-----LGVIPRAIHDI 112
 62 EVVNGAVAPLIKGLFKGYNATVLAGYGTGSGKTYSGGAYTHNQENPTVGVIPRIVIAL 121
 113 FQKIKKFPDRELLRVSYMEIYNETITDLL-CGTQKMKPLIIRREDYNNVYVADLTVEEV 171
 122 FREIHORPEWERNLKVSYLEIYNEEILDLLIYAARDKTNWISIREDPKEGKICGLTERDV 181
 172 YTESEALKWITKEGSRHYGETKONORSRSHYTFRMILESREKSGPSNCEGSKVYSHLN 231
 182 KTAIDTLSCLEQNSRSTVASTAMNSOSSRSHAFITISIEORKEGDKNN--SFR-SKLH 237
 232 LVDLAGSERAAQTGAAGVRLKEGCNINRSFILGQVKKLSD-GVGGFYNYRDSKLTTRI 290
 238 LVDLAGSEROKTKAEGDKLKEGINSRGLLGNVLSALGDESKKGFFVYRDSKLTREL 297
 291 LQNSLGNPKTRIITCTTPV--SFDETLTALQFASAKYMKNTPYVNEVSTDEALLKRYR 348
 298 LQDSLGNSSHTLMIACVSPADSNWETINTLYRADRKIKNKPVN-TDPQAAELQRLK 356
 349 KEIMDLKKQLEEV-----SLETRAQAMEKD-----OLAQ 377
 357 LQVQELQVLLQAHGGTLPVLSNMEPSNLQSLMERNKLEKNGKLSRELGEAAVQ 416
 378 LLEEKOLLQKQNEKI-----ENLTRLVLT 402
 417 FL-EKIIMTEQNEKLGSKRMELKQHAACKVNUQLRVET 454

RESULT 13
 KF17 MOUSE
 ID KF17 MOUSE STANDARD; PRT; 1038 AA.
 AC Q99PW8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Kinesin-like protein KIF17 (MmKIF17).
 GN KIF17.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20307907; PubMed=10846156;
 RA Setou M., Nakagawa T., Seog D.-H., Hirokawa N.;
 RT "Kinesin superfamily motor protein Kif17 and mLin-10 in NMDA
 receptor-containing vesicle transport.";
 RL Science 288:1796-1802(2000).
 CC -!- FUNCTION: Transports vesicles containing N-methyl-D-aspartate
 (NMDA) receptor 2B along microtubules.
 CC -!- SUBUNIT: Interacts with LIN-10 PDZ domain.
 CC -!- TISSUE SPECIFICITY: Neuronal-specific.
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

291 LQNSLGNPKTRIITCTTPV--SFDETLTALQFASAKYMKNTPYVNEVSTDEALLKRYR 348
 296 LQDSLGNPKTRIITCTTPV--SFDETLTALQFASAKYMKNTPYVNEVSTDEALLKRYR 354
 349 KEIMDLKKQLEEV-----SLETRAQAMEKD-----OLAQ 377
 355 BEIKKLKALL-----TQKSPSLSALLSROVPPDPVQVEEKLKLPQVLIQHOME 403

RESULT 12
 KF4A XENLA
 ID KF4A XENLA STANDARD; PRT; 1226 AA.
 AC Q91784; Q9PSI0;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chromosome-associated kinesin KLP1 (Chromokinesin).
 GN KLP1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
 RC SPECIFICITY.
 RX TISSUE=Oocyte;
 RX MEDLINE=95236444; PubMed=7720067;
 RA Vernos I., Raats J., Hirano T., Heasman J., Karsenti E., Wylie C.;
 RT "Xklp1, a chromosomal Xenopus kinesin-like protein essential for
 spindle organization and chromosome positioning.";
 RL Cell 81:117-127(1995).
 RN [2]
 RP SEQUENCE OF 9-338 FROM N.A.
 RX MEDLINE=93246065; PubMed=8482413;
 RA Vernos I., Heasman J., Wylie C.;
 RT "Multiple kinesin-like transcripts in Xenopus oocytes.";
 RL Dev. Biol. 157:232-239(1993).
 CC -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR
 SPINDLE STABILIZATION.
 CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic
 chromosomes.
 CC -!- TISSUE SPECIFICITY: Expressed in oocytes, eggs, testes and
 brain.
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
 CC Chromokinesin subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; X82012; CAA57539.1; --
 F1R; I51617; I51617.
 DR HSP; P17119; 3KAR.
 DR InterPro; IPR001752; kinesin motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; DNA-binding;
 KW Nuclear protein; Coiled coil.
 FT DOMAIN 1 350 KINESIN-MOTOR.
 FT DOMAIN 351 1006 COILED COIL (BY SIMILARITY).
 FT DOMAIN 1007 1226 GLOBULAR.
 FT NP BIND 87 94 ATP (POTENTIAL).
 FT CONFLICT 163 163 I -> L (IN REF. 2).
 FT SEQUENCE 1226 AA; 138923 MW; 7F0275FCF3316697 CRC64;

```

DR EMBL; AB008967; BAB21099.1; -.
DR HSP; P17119; 3KAR.
DR MGD; MGI:1098229; Kif17.
DR GO; GO:0005871; C:kinesin complex; IDA.
DR GO; GO:0003777; F:Microtubule motor activity; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0007017; P:Microtubule-based process; IDA.
DR GO; GO:0016192; P:vesicle-mediated transport; IDA.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN; 1.
DR PROSITE; PS0067; KINESIN MOTOR DOMAIN; 1.
KW Motor protein; Microtubule; ATP-binding; Coiled coil; Transport;
KW Protein transport.
FT DOMAIN 1 265 KINESIN-MOTOR.
FT DOMAIN 346 470 COILED COIL (POTENTIAL).
FT DOMAIN 748 855 COILED COIL (POTENTIAL).
FT NP_BIND 91 98 ATP (POTENTIAL).
SQ SEQUENCE 1038 AA; 116372 MW; 2BED852A3AFDBD46 CRC64;

Query Match 33.3%; Score 683; DB 1; Length 1038;
Best Local Similarity 41.3%; Pred. No. 4e-36;
Matches 166; Conservative 66; Mismatches 140; Indels 30; Gaps 10;

QY 5 AVAVCVVRPLNSRESLG-ETAQVYWKTDNNVYQVDSG-----KSFNFRVFGHNETT 58
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5 SVKVVRCPMKRERELSCQSVTVDSARGQCQFQNGAADPPKQFTFDGAYIEHT 64
QY 59 KNVYEIAPIIDSALQNGNGTIFAYGTASGKTYTMGSED---HLGVIPRAIHDF 115
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
65 EQIYNEIAYPLVEGTEGNGTIFAYGTGSGKFTMQGLPDPQGRGIIPRAEFHVES 124
QY 116 IKFPDFRFLVSYMEIYNETITDLCGTQKMKPLIPREDVNRNYYADLITEEVVY 175
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 VQCAENTFLVRSVLEIYNEVDHLL-GADTKQKLEKEHPEKGVYVYKGLSMHTV 183
QY 176 MALKWITKGEKSRHYGETKMNORSRSHITFRMILESRKEGEPNCEGSKVSHNL 232
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 QCEVWETGWKNRAGVYGLMKNDSKSHSIFINTIEIYAVDERG-----KQHLRAGK 238
QY 233 VDLASERAQGTGAAGVRLKEGCNINRSLFILGVKIKLSQGQGGFINYRDSKLT 292
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
239 VDLASGERQSKTGATGERLKEATKINLSLGNVISALVDGRC-KHIPYRDSKLT 297
QY 293 NSLGNPKTRIICTTPV--SFDETLTALQASTAKYMKNTPYNEVSTDEALLKRYR 350
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 DSLGNTKTLVACLSPADNNYDETLSTLRANRKNINKNKPERINEDPKD-ALLREY 356
QY 351 IMDLKKOLEVSLBETRAQAMEKDQLAQLLEBKDLLQKVNQEK 392
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
357 IKRLKAIL-----AQQMGPNLSALLSTQTPPGVQSEE 390

RESULT 14
KF4A_CHICK
ID KF4A_CHICK STANDARD; PRT; 1225 AA.
AC Q90640; Q90608;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chromosome-associated kinesin KIF4A (Chromokinesin).
GN Kif4A.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
SPECIFICITY.
RC STRAIN=White leghorn; TISSUE=Embryonic retina;
RX MEDLINE=95181533; PubMed=7876303;
RA Wang S.Z., Adler R.;
RT "Chromokinesin: a DNA-binding, kinesin-like nuclear protein.";
RL J. Cell Biol. 128:761-768 (1995).
[2]
RN RP SEQUENCE OF 728-1088 FROM N.A.
RC STRAIN=White leghorn; TISSUE=Embryonic retina;
RX MEDLINE=94151328; PubMed=8108415;
RA Wang S.Z., Adler R.;
RT "A developmentally regulated basic-leucine zipper-like gene and its
expression in embryonic retina and lens.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:1351-1355 (1994).
CC -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR
SPINDLE STABILIZATION.
CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic
chromosomes.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN PROLIFERATING CELLS;
NEUROEPITHELIUM OF EMBRYOS.
CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
Chromokinesin subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U18309; AAC59666.1; -.
DR EMBL; U04821; AAL18960.1; -.
DR PIR; A56514; A56514.
DR HSP; P17119; 3KAR.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN; 1.
DR PROSITE; PS0067; KINESIN MOTOR DOMAIN; 1.
KW Motor protein; Microtubule; ATP-binding; DNA-binding;
Nuclear protein; Coiled coil.
FT DOMAIN 1 351 KINESIN-MOTOR.
FT DOMAIN 352 1003 COILED COIL (BY SIMILARITY).
FT DOMAIN 1004 1225 GLOBULAR.
FT NP_BIND 88 95 ATP (POTENTIAL).
FT CONFLICT 1087 1088 KG -> RI (IN REF. 2).
SQ SEQUENCE 1225 AA; 138923 MW; FA01ED83425F5875 CRC64;

Query Match 32.5%; Score 667; DB 1; Length 1225;
Best Local Similarity 39.1%; Pred. No. 5.3e-35;
Matches 168; Conservative 76; Mismatches 148; Indels 38; Gaps 13;

QY 2 EGVAVAVCVVRPLNSRESLG-ETAQVYWKTDNNVYQVDSGSKSFNFRVFGHNETTKN 60
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6 EKGIPVVRVRCRPLVPKTEGQCMCLSFVPGEQVFI--VGSDFKFTYDYVDFSVSE 63
QY 61 VYIEIAPIIDSALQNGNGTIFAYGTASGKTYTMG-----SEDH---LGVIPRAIHDF 113
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
64 VFTAVAPLIRGIFKGYNAIVLAYGQTGSKTYSMGQTYTASQEHDPMSGVIVPVKILF 123
QY 114 QKIKKPPDRFLRVSYMEIYNETITDLCGT-QKMKPLIREDVNRNYYADLITEEVVY 172
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
124 KEKQKQDWEFLVRSVLEIYNEVDHLLDLCSSRSRSQISREDPKGKIVGLTERNVA 183
QY 173 TSMALKWITKGEKSRHYGETKMNORSRSHITFRMILESRKEGEPNCEGSKVSHNL 232
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 SARDTVSCLEQGNKCTVASTANSSSRSHALFTICIDOKKK---NDKNSFSFKLH 239
QY 233 VDLASERAQGTGAAGVRLKEGCNINRSLFILGVKIKLSQGQGGFINYRDSKLT 291
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
240 VDLASGERQKTKAEGDRLKEGININRGLCLGNVISALGEENKKGFGVYRDSKLT 299
QY 292 QNSLGNPKTRIICTTPV--SFDETLTALQASTAKYMKNTPYNEVSTDEALLKRYR 349

```

Db 300 QDSIGNSHTLMACVSPADSNLETLNLRVADRARKKNPIVN-VDPOAELNHLKQ 358
 QY 350 EIMDLKKQLEE-----VSLTRAQA-----MKDQLAQLLEKDLLOK-----VQNEK 392
 Db 359 QVQQLQVLLQAHGGTLPVSIINSMAPSENLSILMEKNQ--SLMEENEKLSRGLSRAAGT 416
 QY 393 IENLIRMLVT 402
 Db 417 AQMLERIIVT 426

RESULT 15
 KL61_DROME STANDARD; PRT; 1066 AA.
 ID KL61_DROME Q9W019;
 AC P46863; Q8T0A6; Q9W019;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Bipolar kinesin KRP-130 (Kinesin-like protein Klp61F).
 GN KLP61F OR KLP2 OR CG9191.
 OS Drosophila melanogaster [fruit fly].
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephyarodea; Drosophilidae; Drosophila.
 CX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20196006; PubMed=10731132;
 RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fessler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai X.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Riden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Shue B.C., Sengen-Kiamos I., Simpson M., Strong R., Sun E.,
 RA Spier E., Spradling A.C., Stapleton M., Turner E., Wang A.H., Wang X.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S.-C., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP REVISIONS.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley; TISSUE=Embryo;
 RX MEDLINE=22426086; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarin H., Kronmiller B., Paclab J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celniker S.E.;
 RT "A Drosophila full-length cDNA resource.";
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 RN [5]
 RP SEQUENCE OF 228-357 FROM N.A.
 RC STRAIN=DP CN BW;
 RX MEDLINE=92020874; PubMed=1924306;
 RA Stewart R.J., Pesavento P.A., Woerpel D.N., Goldstein L.S.B.;
 RT "Identification and partial characterization of six members of the
 RT kinesin superfamily in Drosophila.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:8470-8474(1991).
 RN [6]
 RP PARTIAL SEQUENCE, AND IDENTIFICATION AS KRP-130.
 RX MEDLINE=97078747; PubMed=8918872;
 RA Kashina A.S., Scholey J.M., Leszyk J.D., Saxton W.M.;
 RT "An essential bipolar mitotic motor.";
 RL Nature 384:225-225(1996).
 CC -!- FUNCTION: IMPORTANT ROLE IN MITOTIC DIVIDING CELLS. MICROTUBULE
 CC MOTOR REQUIRED FOR SPINDLE BODY SEPARATION. SLOW PLUS-END DIRECTED
 CC MICROTUBULE MOTOR CAPABLE OF CROSS-LINKING AND SLIDING APART
 CC ANTIPARALLEL MICROTUBULES. THEREBY PUSHING APART THE ASSOCIATED
 CC SPINDLE FOLDS DURING SPINDLE ASSEMBLY AND FUNCTION.
 CC -!- SUBUNIT: Homotetramer.
 CC -!- DEVELOPMENTAL STAGE: SPECIFICALLY EXPRESSED IN PROLIFERATING
 CC TISSUES DURING EMBRYONIC AND LARVAL DEVELOPMENT.
 CC -!- PTM: PHOSPHORYLATION DURING MITOSIS AT THR-933 CONTROLS THE
 CC ASSOCIATION OF KLP61F WITH THE SPINDLE APPARATUS (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BIMC
 CC SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; U01842; AAA03718.1; -.
 CC EMBL; AB003471; AAF47458.2; -.
 CC EMBL; AY069442; AAL39587.1; -.
 CC EMBL; M74428; AAA28655.1; -.
 CC PIR; A48669; A48669.
 CC HSP; P17119; 3KAR.
 CC FlyBase; FBgn004378; Klp61F.
 CC GO; GO:0005737; C:cytoplasm; IDA.
 CC GO; GO:0008874; C:kinesin complex; IDA.
 CC GO; GO:0003774; F:motor activity; IDA.
 CC GO; GO:0007100; P:centrosome separation; IGI.
 CC InterPro; IPR001752; kinesin_motor.

Search completed: July 29, 2004, 09:36:15
Job time : 4.88047 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:27:31 ; Search time 17.1603 Seconds

(without alignments)

7428.155 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_405

Perfect score: 2053

Sequence: 1 AEGAVAVCVRVPLNSREE.....LQKVQNEKIENLRMLVTSS 404

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1799	87.6	549	11	Q7TPX4 mus musculus
2	1468.5	71.5	2954	13	Q42263
3	872	42.5	807	10	Q94HV9 xenopus lae
4	872	42.5	823	10	Q9S7P3 Q94HV9 arabidopsis
5	856	41.7	459	10	Q9S330 Q9S7P3 arabidopsis
6	836.5	40.7	888	10	Q9L062 Q9S630 arabidopsis
7	784	38.2	1885	5	Q869B8 Q9L062 arabidopsis
8	758.5	36.9	1055	10	Q8RW4 Q869B8 dictyosteli
9	758.5	36.9	1055	10	Q8W5R5 Q8RW4 arabidopsis
10	754	36.7	1459	3	P87198 Q8W5R5 arabidopsis
11	754	36.7	2013	5	Q9VK10 P87198 arabidopsis
12	754	36.7	2244	5	Q9NCG0 Q9VK10 drosophila
13	748.5	36.5	1033	10	Q9LHL9 Q9NCG0 arabidopsis
14	748.5	36.5	1058	10	Q9S7U0 Q9LHL9 arabidopsis
15	727.5	35.4	956	10	Q9C7B9 Q9S7U0 arabidopsis
16	726	35.4	443	11	Q8CGJ1 Q9C7B9 mus musculus

17	724	35.3	997	10	Q9FG03
18	723	35.2	701	11	Q7TSZ7
19	722.5	35.2	959	10	Q8S950
20	721.5	35.1	890	10	Q8W5R6
21	721	35.1	699	5	Q8MPT8
22	720	35.1	408	11	Q8OUK1
23	716.5	34.9	408	4	Q8IWH8
24	715.5	34.9	954	10	Q9AMW8
25	714.5	34.8	599	13	Q98T11
26	713	34.7	747	11	Q8BNH4
27	713	34.7	757	11	Q8OU27
28	707	34.4	160	11	Q3S059
29	706.5	34.4	1193	10	Q7X7H8
30	704	34.3	677	5	Q9VRK9
31	704	34.3	782	5	Q19633
32	704	34.3	782	5	Q9GV92
33	703	34.2	735	5	Q9U0D5
34	702	34.2	974	10	Q8S905
35	701	34.1	836	5	Q7YUC7
36	695	33.9	1037	10	Q9FH58
37	694.5	33.8	671	4	Q86XX7
38	694.5	33.8	1127	4	Q86TN3
39	694	33.8	1121	10	Q9SVI8
40	693	33.8	819	11	Q7TQ66
41	693	33.8	1231	11	Q80YP3
42	692	33.7	909	10	Q9XI03
43	687.5	33.5	727	3	Q86Z96
44	686.5	33.4	581	10	Q9SJU7
45	686	33.4	671	5	Q8MPT7

ALIGNMENTS

RESULT 1

Q7TPX4	PRELIMINARY;	PRT;	549 AA.
ID	Q7TPX4		
AC	Q7TPX4;		
DT	01-OCT-2003 (TRENBLrel. 25, Created)		
DT	01-OCT-2003 (TRENBLrel. 25, Last sequence update)		
DT	01-OCT-2003 (TRENBLrel. 25, Last annotation update)		
DE	Hypothetical protein (Fragment).		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Egg;		
RX	MEDLINE=22388257; PubMed=12477932;		
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,		
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,		
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,		
RA	Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,		
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,		
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,		
RA	Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,		
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,		
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,		
RA	Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,		
RT	"Generation and initial analysis of more than 15,000 full-length human		
RT	and mouse cDNA sequences."		
RT	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Egg;		

RA Strausberg R.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 KW EMBL; BC052843; AAHS2843.1; --
 DR Hypothetical protein.
 FT NON_TER 549 549
 SQ SEQUENCE 549 AA; 62899 MW; B707C97DC9FB38D3 CRC64;
 Query Match 87.6%; Score 1799; DB 11; Length 549;
 Best Local Similarity 87.1%; Pred. No. 2.9e-116; Indels 0; Gaps 0;
 Matches 352; Conservative 17; Mismatches 35;
 QY 1 AEEGAVAVCVVRPLNSREESLGETAQVYKTKDNNVIYQVDSKSFNFRVPHGNETTNN 60
 Db 2 AEEASVAVCVVRPLNSREELGEATHIYKTKDKNALYQSDGKSFQDFRVFDSNETTKN 61
 QY 61 VVEEIAAPIIDSAIQYNGTIFAYGQTASGKTYTMGSEDHLGVIPRAIHDFQIKKFP 120
 Db 62 VVEEIAVPIISSAIQYNGTIFAYGQTASGKTYTMGSEDCLGVIPRAIHDFQIKKFP 121
 QY 121 DREFLARVSMETYNITIDLLCGTOKMKPLIREDVNRVYVADLTVEEVYVYSEMAK 180
 Db 122 EREFLARVSMETYNITIDLLCGTOKMKPLIREDVNRVYVADLTVEEVYVYSEMAK 181
 QY 181 ITKGEKSRHYGKTKMQRSSRSHTIFRMILESEKGEPSNCEGSKVSHNLVLDLAGSER 240
 Db 182 LATGEKRNHYGKTKMQRSSRSHTIFRMILESEKGEPSNCEGSKVSHNLVLDLAGSER 241
 QY 241 AAGTGAAGVRLKGCNCNINRSILFQGVIKLSDGVGGFNYRDSKLTILQNSLGGNPK 300
 Db 242 AAGTGAAGVRLKGCNCNINRSILFQGVIKLSDGVGGFNYRDSKLTILQNSLGGNPK 301
 QY 301 TRICTITPVSFDETLTALOPASTAKYKNTPPYNEVSTDEALKRYRKEIMDLKLEE 360
 Db 302 TRICTITPASLDETLTALOPASTAKYKNTPPYNEVSTDEALKRYRKEIMDLKLEE 361
 QY 361 VSLTEPRAQAMEKDLQALKEKLLQKQVNEKIENLRMLVTSS 404
 Db 362 VNTKTRAQAMEKDLQALKEKLLQKQVDEKINNLRMLVTSS 405
 RESULT 2
 ID O42263 PRELIMINARY; PRT; 2954 AA.
 AC O42263
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Kinesin-related protein.
 GN XCENP-E.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98028574; PubMed=9363944;
 RA Wood K.W., Sakowicz R., Goldstein L.S., Cleveland D.W.;
 RT "CENP-E is a plus end-directed kinetochore motor required for
 metaphase chromosome alignment.";
 RL Cell 91:357-366 (1997).
 DR EMBL; AF027728; AAC60300.1; --
 DR PIR; T14156; T14156.
 DR HSP; F33176; 1B2.
 DR GO; GO:0005871; C:kinesin complex; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003774; F:motor activity; IEA.
 DR GO; GO:0007017; P:microtubule-based process; IEA.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINSEINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN2; 1.
 SQ SEQUENCE 807 AA; 91260 MW; FD7CDAD68EA30C28 CRC64;

DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
 SQ SEQUENCE 2954 AA; 339964 MW; 439804ED0E592679 CRC64;
 Query Match 71.5%; Score 1468.5; DB 13; Length 2954;
 Best Local Similarity 70.9%; Pred. No. 2.3e-92; Indels 7; Gaps 3;
 Matches 290; Conservative 52; Mismatches 60;
 QY 1 AEEGAVAVCVVRPLNSREESLGETAQVYKTKDNNVIYQVDSKSFNFRVPHGNETTNN 60
 Db 2 SEGDAVKCVVRPLIQREQ--GDQANLQWKAGNNTISQVDTGKSFNFRVPHGNETTNN 59
 QY 61 VVEEIAAPIIDSAIQYNGTIFAYGQTASGKTYTMGSEDHLGVIPRAIHDFQIKKFP 120
 Db 60 IQEETAVPIIRALQYNGTIFAYGQTSSGKTYTMGTENSLGIIIPQAIQEVFKIIQEI 119
 QY 121 DREFLARVSMETYNITIDLLCGTOKMKPLIREDVNRVYVADLTVEEVYVYSEMAK 180
 Db 120 NREFLARVSMETYNITIDLLCGTOKMKPLIREDVNRVYVADLTVEEVYVYSEMAK 179
 QY 181 ITKGEKSRHYGKTKMQRSSRSHTIFRMILESEKGEPSNCEGSKVSHNLVLDLAG 237
 Db 180 IKKGEKRNHYGKTKMQRSSRSHTIFRMILESEKGEPSNCEGSKVSHNLVLDLAG 239
 QY 238 SERAAGTGAAGVRLKGCNCNINRSILFQGVIKLSDGVGGFNYRDSKLTILQNSLGG 297
 Db 240 SERAAGTGAAGVRLKGCNCNINRSILFQGVIKLSDGVGGFNYRDSKLTILQNSLGG 299
 QY 298 NPXTRICTITPVSFDETLTALOPASTAKYKNTPPYNEVSTDEALKRYRKEIMDLK 357
 Db 300 NAKTVIICITTPVSFDETLTALOPASTAKYKNTPPYNEVSTDEALKRYRKEIMDLK 359
 QY 358 LE--EVSLEPRAQAMEKDLQALKEKLLQKQVNEKIENLRMLVTSS 404
 Db 360 LENLESSETKAQAMAKBEHTQLLAETKQLKHEREDRIWHLTNIIVASS 408
 RESULT 3
 ID Q94HV9 PRELIMINARY; PRT; 807 AA.
 AC Q94HV9
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Kinesin motor protein (kin2), putative.
 GN T4M14.11.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Town C.D., Haas B.J., Wu D., Maiti R., Hannick L.I., Chan A.P.,
 RA Tallon L.J., Rooney T., Utterback T.R., Vanaken S.E., Feildblyum T.V.,
 RA White O., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome 1 BAC T4M14 genomic sequence.";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC027036; AAK62792.1; --
 DR GO; GO:0005871; C:kinesin complex; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003774; F:motor activity; IEA.
 DR GO; GO:0007017; P:microtubule-based process; IEA.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINSEINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
 SQ SEQUENCE 807 AA; 91260 MW; FD7CDAD68EA30C28 CRC64;
 Query Match 42.5%; Score 872; DB 10; Length 807;
 Best Local Similarity 47.6%; Pred. No. 7.6e-52;


```

RX MEDLINE=98359834; PubMed=9693369;
RA de Hostos E.L., McCaffrey G., Sugang R., Pierce D.W., Vale R.D.;
RT "A developmentally regulated kinesin-related motor proteins from
RT Dictyostelium discoideum.";
RL Mol. Biol. Cell 9:2093-2106(1998).
RN [2]
RN SEQUENCE FROM N.A.
RP Suyama E., Sutoh K.;
RA "Kinesin-related proteins from Dictyostelium.";
RT Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AB102780; BAC56912.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005875; C:microtubule associated complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0001752; kinesin_motor.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR Pfam; PF02370; M; 6.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.
SQ SEQUENCE 1885 AA; 218216 MW; 15686ABD3B007EC7 CRC64;

Query Match 38.2%; Score 784; DB 5; Length 1885;
Best Local Similarity 41.4%; Pred. No. 3.1e-45;
Matches 183; Conservative 81; Mismatches 124; Indels 54; Gaps 10;

Qy 2 EEGAVAVCVVRPLNSREESLGETAQYWKTDNNVI-YQVDGSKSFNDRVFNHNETKN 60
Db 19 ELNKKIKVAIRVPLNSRE-LGIDQKIPWSIKDTISLSQNPINFYDVYFGIDSNITD 76
Qy 61 VYEEIAPIIDSALQGYNGTIFAYGQTASGTYTMGSEDLGVIPRAIHDFOKIK-KF 119
Db 77 VYNAIAKSIIVNSLNGINGTIFAYGQTSSGKTFSMRGTESIPGIKLSIKDIFKSDSI 136
Qy 120 PDREFLAVSYMEIYNETITDLCGT-QKMKPLIREDVNRNVVADLTERRVVTSEMA 178
Db 137 LEKDYLLKVSYLEIYNEIKDNLNPTISNKKKIKHEDIYKGVVAVANLKEEIVISPOIF 196
Qy 179 KWIPTGSKSRHYGTQKQNRSSRSHITPRMILESRKEGPNCSGVKSVSHNLVDLAGS 238
Db 197 ALNFGGEHRHIGSTWMDSSSRSHITPRMQIQSTCKQ-----NGTIQMSTLTLDLAGS 251
Qy 239 ERAAQTGAAGVRLKGCNINRSFILGOVIKLSDGQVGGFINYRDSKLTFRILQNSLGN 298
Db 252 ERVSTGAEGVRLKEGTHINKSLMTLSKVISKLSEKTTQOHVPYRDSKLTFRILQSLGN 311
Qy 299 PKTRICTITPVSF--DETTLALQFASHTAKYMKNTPPYVNEVSTDEALLKRYRKEIMDLK 356
Db 312 SKTALLCTITPATTHQESISTLOFAKRAKRVKNTYKINQVADANTMLKKYSEIILELQ 371
Qy 357 QL-----EEVSL-----ETRAQAMEKQDL--AQ 377
Db 372 QLVKSEERINSLRNTISQEISSNNFKLGMRFNDAIIGSSLINENKKKKRRNTLDPSY 431
Qy 378 LLEEKDLLQKV---QNEKIENL 396
Db 432 LLKDKIINKKIRKGENQKIKKI 453

RESULT 8
Q8RW4 PRELIMINARY; PRT; 1055 AA.
AC Q8RW4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative kinesin.
GN Arabidopsis thaliana (Mouse-ear cress).
OS Arabidopsis thaliana (Mouse-ear cress); Embryophyta; Tracheophyta;
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OX eurosids II; Brassicales; Brassicaceae; Arabidopsais.
RN [1]
RN SEQUENCE FROM N.A.
RP Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
RA Davis R.W., Ecker J.R., Theologis A.;
RT "Arabidopsis Full Length cDNA Clones.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY091060; AM13891.1; -.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IPR001752; kinesin_motor.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.
SQ SEQUENCE 1055 AA; 116462 MW; BBF3709ACB40215B CRC64;

Query Match 36.9%; Score 758.5; DB 10; Length 1055;
Best Local Similarity 42.0%; Pred. No. 8e-44;
Matches 176; Conservative 84; Mismatches 136; Indels 23; Gaps 8;

Qy 1 AEGAVAVCVVRPLNSREESLGETAQYWKTDNNVIYQVDGSKSFNDRVFNHNETKN 60
Db 94 SERDSISVTVRFPLSDREYQRGDEVAWYPDGDTLVRHEYNPLTAYAFDKVFGPQAITD 153
Qy 61 VYEEIAPIIDSALQGYNGTIFAYGQTASGTYTMGSEDLGVIPRAIHDFOKIK-KF 120
Db 154 VYDVAAPVVKAAAMEGVNGTVFAYGVTSSGKTHTMHGDQESPGHIPLAKDVFSIQDTP 213
Qy 121 DREPLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVVADLTERRVVTSEMA 180
Db 214 GRELLAVSYLEIYNEIYNEINDLLDPTG--QNLRVRED--SQGYVEGKEEVVLSPGHALSF 270
Qy 181 ITGGEKSRHYGTQKQNRSSRSHITPRMILESRKEGPNCSGVKSVSHNLVDLAGSER 240
Db 271 IAAGEEHRHVGSNFNLLSSRSHTIFTLMVESATGDEYD---GVIFSQLNLDLAGSE- 326
Qy 241 AAQTGAAGVRLKGCNINRSFILGOVIKLSDGQVGGFINYRDSKLTFRILQNSLGNPK 300
Db 327 SSKTETTLRRKKEGYNKSLTLTGTVIGKLSGK-ATHIPYRDSKLTFRILQSSLSGHG 385
Qy 301 TRIICTITPV--SPDETTLALQFASHTAKYMKNTPPYVNEVSTDEALLKRYRKEIMDLK 358
Db 386 VSLICTITPASSSEETHNTLKFSRAKSIETIYASRNQIIDEKSLIKKYQREISTLKLEL 445
Qy 359 EEV-----SLETRAQAMEKQD--LAQLLEKDLLQKVQNEKIENITRMLVTSS 404
Db 446 DQLRRGMLGVGVSHBELMSLKQLEEGQVQMGQSRLEEBEAKAALMSRIQKLTKLIVST 504

RESULT 9
Q8W5R5 PRELIMINARY; PRT; 1055 AA.
AC Q8W5R5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Kinesin-related protein (Putative kinesin protein).
GN MKRP2 OR AT4G39050.

```

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopses.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21563048; PubMed=11706156;
 RA Itoh R., Fujiwara M., Yoshida S.;
 RT "Kinesin-Related Proteins with a Mitochondrial Targeting Signal";
 RL Plant Physiol. 127:724-726(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
 Quach H.L., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
 Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,
 Palm C.J., Shinn P., Southwick A., Tripp M.G., Wu T., Davis R.W.,
 Ecker J.R., Theologis A.;
 RT "Arabidopsis Open Reading Frame (ORF) Clones";
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB062739; BAB71852.1; -.
 DR EMBL; AX150516; AAN13032.1; -.
 DR GO; GO:0005871; C:kinesin complex; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003774; F:motor activity; IEA.
 DR GO; GO:0007017; P:microtubule-based process; IEA.
 DR InterPro; IPR001752; P:microtubule-based process; IEA.
 DR InterPro; IPR001841; Znf ring.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR SMART; SM00184; RING; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
 DR PROSITE; PS0089; ZF RING 2; 1.
 SQ SEQUENCE 1055 AA; 116463 MW; B69EC383FF14AB7B CRC64;

Query Match 36.9%; Score 758.5; DB 10; Length 1055;
 Best Local Similarity 42.0%; Pred. No. 8e-44;
 Matches 176; Conservative 84; Mismatches 136; Indels 23; Gaps 8;

QY 1 AEEGAVAVRVRLNRSREESLGETAQVYWKTDN--NVIYQVD-----GSKSFNDRVPHGNETKNVVEETAAPIIDSAIOGYNGTIFAYGQ 60
 DB 94 SERDSIVTVRPLSDREYQRGDEVAVTPDGLVRHEYNPLTAYAFKDPGPQATTID 153
 QY 61 VYEEIAAPIIDSAIOGYNGTIFAYGOTASGKTYMMGSEDHGLGVI PRAIHDFIQIKKFP 120
 DB 154 VYDVAARPVVKAAEGVGTGTFAYGVTSSGKTHTMHGDDQESPGLIPLAIKDNVESIIQDTP 213
 QY 121 DREFLRVSYMEYNETIIDLLCGTQKMKPLIIRVDNVRNVYVADLTVEVYVTSSEALKW 180
 DB 214 GREFLRVSYLETYNEVINDLDPTG--QNLVRVED--SQGTVEGKEEVLSPGHALSF 270
 QY 181 ITKGEKSRHYGETKMNORSRSHITFRMILESREKGEPCNSGVKSHNLNVDLAGSER 240
 DB 271 IAGEHRHVGNFNFLNLSRSHITFTLVVSSATGDEYD---GVTFPSQNLNLDLAGSE- 326
 QY 241 AAQTGAAGVRLKEGCNINRSLFILGQVIKKLSQGVGFNYRDSKLITRILQNSLGNPK 300
 DB 327 SSKTETTLRRKEGSYINKSLTLGTGVIGKLSGK-ATHIPYRDSKLITRILQNSLGHG 385
 QY 301 TRLICITIPV--SFDETLALQFASAKYKNTPTPYNEVSTDEALKRYRKEIMDLKQL 358
 DB 386 VSLICTITPASSSEETHNTLPASRAKSIETIYASRNQIIDEKSLIKKYQREISTUKLEL 445
 QY 359 EEV-----SLETRAAMEKQD--LAQLEEKDLQKQVNEKIENTRLMVLTS 404
 DB 446 DQLRRGMVGVSHLEMLSLKQLEEGQVKKQSELEEEBAKALMSRIKILKLIIVST 504

RESULT 10

P87198

ID P87198 PRELIMINARY; PRT; 1459 AA.

AC P87198;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Kinesin motor protein.
 GN KIN1.
 OS Ustilago maydis (Smut fungus).
 OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
 OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
 OX NCBI_TaxID=5270;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97361828; PubMed=9218789;
 RA Lehmler C., Steinberg G., Snetselaar K.M., Schliwa M., Kahmann R.,
 Bolker M.;
 RT "Identification of a motor protein required for filamentous growth in
 Ustilago maydis";
 RL EMBO J. 16:3464-3473(1997).
 DR EMBL; U28844; AAB63336.1; -.
 DR PIR; T30196; T30196.
 DR HSSP; P17119; 3KAR.
 DR GO; GO:0005871; C:kinesin complex; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003774; F:motor activity; IEA.
 DR GO; GO:0007017; P:microtubule-based process; IEA.
 DR InterPro; IPR001752; kinesin; 1.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
 SQ SEQUENCE 1459 AA; 156932 MW; 281B74B0100F5CDO CRC64;

Query Match 36.7%; Score 754; DB 3; Length 1459;
 Best Local Similarity 38.2%; Pred. No. 2.6e-43;
 Matches 190; Conservative 70; Mismatches 127; Indels 110; Gaps 14;

QY 6 VAVCVRVRLNRSREESLGETAQVYWKTDN--NVIYQVD-----GSKSFNDRVPHGNETKNVVEETAAPIIDSAIOGYNGTIFAYGQ 86
 DB 237 VVVCVRMRP--SRASSDSEASV-WNCDSKRNIFPTEHHPALAKRTTSERAGASIA 293
 QY 42 -----GSKSFNDRVPHGNETKNVVEETAAPIIDSAIOGYNGTIFAYGQ 353
 DB 294 AAPSSHDLHDHEDPTSYTHYFQDKLITGAQTDDMYHSHIAPVVRAAVEGNGTIFAYGQ 353
 QY 87 TAGSKTYTMMGSEDHGLGVI PRAIHDFIQIKKFPDREFLRVSYMEYNETIIDLLCGTQ 146
 DB 354 TSGSKTHTWSSGDAEPGVI PRAVEQIFQMIKDEPDREFLRVSYLEIYNETLKDLA-- 410
 QY 147 KMKPL-----IREDVNRNVYVADLTVEV 171
 DB 411 PLPPLTGTSGSLQTTDRPASPICKGSSHAAGSQSCTLRRIEDQKSSRVIITGLREEIV 470
 QY 172 YTSSEALKWITKGEKSRHYGETKMNORSRSHITFRMILESREKGEPCNSGVKSHNLN 231
 DB 471 TANTVLCUIQRGDRRHVGATDWNERSRSHCVFOLTIESRPPAPSASKE--VRISQLN 528
 QY 232 LVDLAGSERAAQTGAAGVRLKEGCNINRSLFILGQVIKKLSQGVGF--FINYRDSKLTR 289
 DB 529 LIDLAGSERAA--SQAERRKEGAFINKSLTLGTGVIGKLTPEVNGDAHI PYRDSKLTR 585
 QY 290 ILQNSLGNPKYTRICTITPV--SFDETLALQFASAKYKNTPTPYNEVSTDEALKRY 347
 DB 586 ILQTSLSGNARIAVICTLSPDTEHANEITSLTKFGKRCVKLVVTTAKKGTAMDMDKALLQKY 645
 QY 348 RXEIMDLKKOLEE-----VSLERAAMEK--DOLAOLLE-----EKDLLQKVQ 389
 DB 646 RRELDALRAKLKANGSPNPNEVMTIVSAESKESQKLDQLNQKEAAQREVEDMKKR 705
 QY 390 NE---KIENLRLMVLTS 403
 DB 706 SHLKAQIEHLRLILTS 722

Best Local Similarity 45.4%; Pred. No. 4e-43;
Matches 181; Conservative 62; Mismatches 118; Indels 38; Gaps 10;

QY 5 AVAVCVVRPLNSRRESLGCTAQTWYTKT-DNNVIVQVDG-SKSFNDRVFGHNETTKNY 62
ID Q9VK10 PRELIMINARY; PRT; 2013 AA.
AC Q9VK10; DB 8 SIQVCIKVRPCPGTSL-----WQVKRSIHLASHAPYVDFYVVDGASNQEVF 60
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Cmet protein.
DE CMET OR CG6392.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=Berkelley;
RC MEDLINE=20196006; PubMed=10731132;
RQ Adams M.D., Ceinliker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RRA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RRA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RRA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RRA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
RRA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RRA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RRA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RRA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RRA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RRA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RRA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RRA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RRA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RRA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RRA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RRA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RRA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RRA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RRA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RRA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RRA Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RRA Liu X., Mattai B., McIntosh J.C., McLeod M.P., McPherson D.,
RRA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RRA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RRA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleeb J.M.,
RRA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RRA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RRA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RRA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RRA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RRA Wang Z.-Y., Wasmann D.A., Worley K.C., Wu D., Yang S., Yao Q.A.,
RRA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RRA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RRA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RRA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RRA "The genome sequence of Drosophila melanogaster.";
RRA Science 287:2185-2195(2000).
RRE ENBL; AE003631; AAF53088.2; -.
RRL HSP; P33176; I8G2.
RRL FlyBase; FBgn0040232; cmet.
RRL GO; GO:0005699; C:kinetochore; IDA.
RRL GO; GO:0007080; P:mitotic metaphase plate congression; IMP.
RRL InterPro; IPR001752; Kinesin_motor.
RRL InterPro; IPR002017; Spectrin.
RRL Pfam; PF00225; kinesin; 1.
RRL PRINTS; PR00380; KINESINHEAVY.
RRL SMART; SM00129; KISC; 1.
RRL SMART; SM00150; SPEC; 3.
RRL PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
RRL PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
RRL PROSITE; PS50067; KINESIN MOTOR DOMAIN2; 1.
RRL SEQUENCE 2013 AA; 23107 MW; 8A14A65581EF4F5F CRC64;

Query Match 36.7%; Score 754; DB 5; Length 2013;
Best Local Similarity 45.4%; Pred. No. 4.7e-43;
Matches 181; Conservative 62; Mismatches 118; Indels 38; Gaps 10;

QY 5 AVAVCVVRPLNSRRESLGCTAQTWYTKT-DNNVIVQVDG-SKSFNDRVFGHNETTKNY 62
ID Q9NCGO PRELIMINARY; PRT; 2244 AA.
AC Q9NCGO; DB 8 SIQVCIKVRPCPGTSL-----WQVKRSIHLASHAPYVDFYVVDGASNQEVF 60
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Kinesin-like kinetochore motor protein CENP-meta.
DE CMET OR CG6392.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=20351410; PubMed=10893249;
RRA Yucel J.K., Marzalek J.D., McIntosh J.R., Goldstein L.S.B.,
RRA Cleveland D.W., Philp A.V.;
RRA "CENP-meta, an Essential Kinetochores Kinesin Required for the
RRA Maintenance of Metaphase Chromosome Alignment in Drosophila.";
RRL J. Cell Biol. 150:1-12(2000).
RRL ENBL; AF220353; AAF32355.1; -.
RRL HSP; P33176; I8G2.
RRL FlyBase; FBgn0040232; cmet.
RRL GO; GO:0005699; C:kinetochore; IDA.
RRL GO; GO:0007080; P:mitotic metaphase plate congression; IMP.
RRL InterPro; IPR001752; kinesin_motor.
RRL Pfam; PF00225; kinesin; 1.
RRL PRINTS; PR00380; KINESINHEAVY.
RRL SMART; SM00129; KISC; 1.
RRL PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
RRL PROSITE; PS50067; KINESIN MOTOR DOMAIN2; 1.
RRL SEQUENCE 2244 AA; 257992 MW; FAGAA3B2A541ADE0 CRC64;

Query Match 36.7%; Score 754; DB 5; Length 2244;
Best Local Similarity 45.4%; Pred. No. 4.7e-43;
Matches 181; Conservative 62; Mismatches 118; Indels 38; Gaps 10;

QY 5 AVAVCVVRPLNSRRESLGCTAQTWYTKT-DNNVIVQVDG-SKSFNDRVFGHNETTKNY 62
ID Q9NCGO PRELIMINARY; PRT; 2244 AA.
AC Q9NCGO; DB 8 SIQVCIKVRPCPGTSL-----WQVKRSIHLASHAPYVDFYVVDGASNQEVF 60
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Kinesin-like kinetochore motor protein CENP-meta.
DE CMET OR CG6392.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=20351410; PubMed=10893249;
RRA Yucel J.K., Marzalek J.D., McIntosh J.R., Goldstein L.S.B.,
RRA Cleveland D.W., Philp A.V.;
RRA "CENP-meta, an Essential Kinetochores Kinesin Required for the
RRA Maintenance of Metaphase Chromosome Alignment in Drosophila.";
RRL J. Cell Biol. 150:1-12(2000).
RRL ENBL; AF220353; AAF32355.1; -.
RRL HSP; P33176; I8G2.
RRL FlyBase; FBgn0040232; cmet.
RRL GO; GO:0005699; C:kinetochore; IDA.
RRL GO; GO:0007080; P:mitotic metaphase plate congression; IMP.
RRL InterPro; IPR001752; kinesin_motor.
RRL Pfam; PF00225; kinesin; 1.
RRL PRINTS; PR00380; KINESINHEAVY.
RRL SMART; SM00129; KISC; 1.
RRL PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
RRL PROSITE; PS50067; KINESIN MOTOR DOMAIN2; 1.
RRL SEQUENCE 2244 AA; 257992 MW; FAGAA3B2A541ADE0 CRC64;

Query Match 36.7%; Score 754; DB 5; Length 2244;
Best Local Similarity 45.4%; Pred. No. 4.7e-43;
Matches 181; Conservative 62; Mismatches 118; Indels 38; Gaps 10;

QY 5 AVAVCVVRPLNSRRESLGCTAQTWYTKT-DNNVIVQVDG-SKSFNDRVFGHNETTKNY 62
ID Q9NCGO PRELIMINARY; PRT; 2244 AA.
AC Q9NCGO; DB 8 SIQVCIKVRPCPGTSL-----WQVKRSIHLASHAPYVDFYVVDGASNQEVF 60
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Kinesin-like kinetochore motor protein CENP-meta.
DE CMET OR CG6392.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=20351410; PubMed=10893249;
RRA Yucel J.K., Marzalek J.D., McIntosh J.R., Goldstein L.S.B.,
RRA Cleveland D.W., Philp A.V.;
RRA "CENP-meta, an Essential Kinetochores Kinesin Required for the
RRA Maintenance of Met

8 SIQVCIKVRPCPLTSL-----WQVKEGRSIIHLADSHABPVYDFVDEGASQNEVF 60
 63 EIAPIIDSAIQYNGTTFAYQOTASGKTYTMMGSEDHLGVIPRAIHDFQIKKFPDR 122
 61 DRMARHIVACMQGFGTTFAYQOTSSGKTYTMMGDEQNPVWVLAKEIFQOISSETER 120
 123 ELLRVSYMEIYNETITDILGCGQKMKPLIIRDVNRNVYVADLTEEYVVTSEM-ALKWI 181
 121 DFLRVGYIEIYNEKIYDIL--NKKNQDLKIHESGNGVNVN--CEECIITSEVDLLRL 176
 182 TKGEKSRHGETKONQSSRSHTIFRMILESRKGPSPNCEGSKVYSHLNLVDLAGSERA 241
 177 CLNKERTVGETWNERSRSRSHAFKIIIESR-KSDHSD-DDAVIOVLNLVDLAGSERA 234
 242 AQTGAAGVRLKEGNCINRSFILGOVTKKLSGQVGVGFNRYRDSKLTIRLQNSLGGNPKT 301
 235 DOTGARGARLKEGHINKSLFISNVTKSUSENADNFTYRYSKLTIRLQNSLGGNAFT 294
 302 RIICITPPVSFDETLTALQFASAKYMKNTFYVNEVSTDEALLKRYRKEIMDLKKOLEEV 361
 295 SIICITKPSIMESQSLSFATRAKKIRIKPQVNMVSDATMMKRLEREIKVLKDLAE- 353
 362 SLETRAQMBKQDLAQLLEKDLQVQNEKIEHLEML 400
 354 -----EER---KNENQOKVEHLERQI 371

RESULT 13
 ID Q9LHL9 PRELIMINARY; PRT; 1033 AA.
 AC Q9LHL9;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Kinesin (Centromere protein) like heavy chain-like protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=20363099; PubMed=10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT TAC and BAC clones.";
 RL DNA Res. 7:217-221(2000).
 DR HSP; P33176; 1BG2.
 DR GO; GO:0005871; C:kinesin complex; IEA.
 DR GO; GO:0005524; P:ATP binding; IEA.
 DR GO; GO:0003774; F:motor activity; IEA.
 DR GO; GO:0007017; P:microtubule-based process; IEA.
 DR InterPro; IPR001752; P:kinase motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAV.
 DR SMART; SM00129; KTSC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS50067; KINESIN MOTOR DOMAIN2; 1.
 DR SEQUENCE 1033 AA; 116718 MW; 741F5977FD6048FB CRC64;
 Query Match 36.5%; Score 748.5; DB 10; Length 1033;
 Best Local Similarity 41.8%; Pred. No. 3.9e-43;
 Matches 179; Conservative 77; Mismatches 131; Indels 41; Gaps 11;
 6 VAVCVRVRLNSREESLGETAQVYVWKTNNVIVQDGSK--SPNFDRVFGHNETKNVYE 63

70 VVTVVRFLSPREIRQGE--EVAWADGETIVRNEHNPITAYAVDRVFGPTTTTNNVYD 127
 64 EIAPIIDSAIQYNGTTFAYQOTASGKTYTMMGSEDHLGVIPRAIHDFQIKKFPDR 123
 128 IAAHVWVNGAMEGINTTFAYGVTSKGTHTHGDORSFGIIPLAVKDAFSLIOETPNE 187
 124 FLIRVSYMEIYNETITDILGCGQKMKP-----LIREDVNRNVYVADLTEEYVVTSEMALK 179
 188 FLIRVSYMEIYNEVNDLL-----NPAGNHLRIED-KQGTVEGIEKEEVLSAPAHLS 240
 180 WITKGE---KSRHYGETKONQSSRSHTIFRMILESRKGPSPNCEGSKVYSHLNLVDLA 236
 241 LIAAGEVMEIQRHVGSNENLLSSRSHITFTLTISSPLGDKSGE-AVHLSQNLVDLA 299
 237 GSRAAQTGAAGVRLKEGNCINRSFILGOVTKKLSGQVGVGFNRYRDSKLTIRLQNSL 296
 300 GSE-SSKVETSQVRRKEGYSINKSLTLGTIVSKLTDVR-ASHVPFRDSKLTIRLQSSLS 357
 297 GNPKTRITCITPV--SFDETLTALQFASAKYMKNTFYVNEVSTDEALLKRYRKEIMDL 354
 358 GHRVSLICITVTPASSSEETHNTLFAHRAKHIEIQAEQNKIIDEKSLIKYQREIRQL 417
 355 KKOLEEVSLETRAQAMEKD-----QLAQLLEKDLQVQNEKIEHLEML 396
 418 KEELEQLKQBEIVPVPQLKDIGADDIVLLKQLEQGVKQSRLEEEEBEAKALLSRQL 477
 397 TRMLVTSS 404
 478 TKLIVST 485
 RESULT 14
 ID Q9SJU0 PRELIMINARY; PRT; 1058 AA.
 AC Q9SJU0; Q94BQ1;
 DT 01-JUN-2000 (TrEMBLrel. 13, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative kinesin heavy chain.
 GN AT2G21380.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
 RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,
 RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,
 RA Fraser C.M., Venter J.C.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Dale J.M., Gibson H.A., Goldsmith A.D., Jiang P.X., Lee J.M.,
 RA Quach H.L., Tang C., Toriumi M., Yu G., Bowser L., Carninci P.,
 RA Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
 RA Karlin-Neumann G., Kawai J., Kim C., Koeseema E., Lam B., Lin J.,
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
 RA Quach H.L., Tang C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
 RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,

Fri Aug 6 10:49:11 2004

QY 358 LEEVLETRAQAMEYD-----OLAQLLEKDLLOKVONEKINLTRM 399
||:| |
Db 409 LEQLKOEIVVPQLXKDIGADDIVLLKQKLEDCQVKLOSRLEEEERAKALLSRIQRLTKL 468
QY 400 LVTSS 404
Db 469 ILVST 473

Search completed: July 29, 2004, 09:39:41
Job time : 18.1603 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:26:01 ; Search time 27:9291 Seconds
(without alignments)
4694.096 Million cell updates/sec

Title: US-10-045-631b-88_COPY_2_465
Perfect score: 2360
Sequence: 1 AEEGAVAVCVRPVPLNSREE.....ITTKTKLSINLLRIDESV 464

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2360	100.0	2633	4	ABG06505	Abg06505 Novel hum
2	2360	100.0	2663	4	AAM39097	Aam39097 Human pol
3	2251.5	95.4	2688	4	AAM40883	Aam40883 Human pol
4	1549	65.6	2954	2	AAy01632	AAy01632 Amino aci
5	820	34.7	366	4	ABU53125	ABU53125 Intracell
6	781	33.1	348	4	ABU53208	ABU53208 Human cel
7	778.5	33.0	2013	4	ABB62322	ABb62322 Drosophil
8	716.5	30.4	677	4	ABB65183	ABb65183 Drosophil
9	704	29.8	1029	5	AAE17786	AAe17786 Human kin
10	704	29.8	1029	5	ADCl0190	ADc10190 Human NOV
11	697.5	29.6	1232	7	ADd49938	ADd49938 Human lun
12	697.5	29.6	1232	7	ADd18924	ADd18924 Human dis
13	695	29.4	473	5	ABG70992	ABg70992 Human tar
14	695	29.4	522	5	ABG70991	ABg70991 Human HSK
15	695	29.4	1038	5	AAM48337	Aam48337 Human kin
16	694.5	29.4	1048	4	ABBS9245	ABb59245 Drosophil
17	694.5	29.4	1066	6	ABG72693	ABg72693 Fruitfly
18	694.5	29.4	1066	6	ABG72693	ABg72693 Fruitfly
19	694	29.4	1034	3	AAG31112	Aag31112 Arabidops
20	694	29.4	1089	3	AAG31111	Aag31111 Arabidops
21	694	29.4	1121	3	AAG31110	Aag31110 Arabidops
22	688.5	29.2	1234	5	ABG70993	ABg70993 Human HSK
23	686.5	29.1	1232	7	ADd49937	ADd49937 Human kin
24	686.5	29.1	1232	7	ADd49932	ADd49932 Human lun
25	680.5	28.8	1232	5	ABG70990	ABg70990 Human HSK

ALIGNMENTS

RESULT 1
ABG06505
ID ABG06505 standard; protein; 2633 AA.

XX AC ABG06505;

DT DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #6496.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX FN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX FI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX N-PSDB; AAS70692.

XX PT New isolated polynucleotide and encoded polypeptides, useful in

XX PT diagnostics, forensic, gene mapping, identification of mutations

XX PT responsible for genetic disorders or other traits and to assess

XX PT biodiversity.

XX PS Claim 20; SEQ ID NO 36864; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)

XX CC sequences. (I) is useful as hybridisation probes, polymerase chain

XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

XX CC and in recombinant production of (II). The polynucleotides are also used

XX CC in diagnostics as expressed sequence tags for identifying expressed

XX CC genes. (I) is useful in gene therapy techniques to restore normal

Adb67093 Kinesin h
Aau19569 Human dia
Abp51294 Human MDD
Ade55349 Rat Prote
Aaw72746 Drosophil
Abb63485 Drosophil
Adb67088 Kinesin h
Aaw72745 Drosophil
Aaw72744 Drosophil
Aau74840 Human HSK
Aae14400 Human kin
Aau79590 Human kin
Adb48222 Human bla
Adb80468 Ovarian c
Adc35116 Human bre
Abp68930 Human pol
Abg60124 Human DIT
Abb2962 Drosophil
Adb67091 Kinesin h
Aag31117 Arabidops

26 664.5 28.2 1031 7 ADB67093
27 658 27.9 757 4 AAU19569
28 658 27.9 757 5 ABP51294
29 656 27.8 796 7 ADE55349
30 655.5 27.8 975 2 AAW72746
31 655.5 27.8 975 4 ABB63485
32 655.5 27.8 975 7 ADB67088
33 652 27.6 411 2 AAW72745
34 652 27.6 441 2 AAW72744
35 652 27.6 1362 5 AAU74840
36 652 27.6 1388 5 AAE14400
37 652 27.6 1388 5 AAU79590
38 652 27.6 1388 6 ADB48222
39 652 27.6 1388 7 ADB80468
40 652 27.6 1388 7 ADB80468
41 652 27.6 1805 5 ABP68930
42 651 27.6 762 5 ABB60124
43 650 27.5 1921 4 ABB62962
44 648.5 27.5 963 7 ADB67091
45 647.5 27.4 829 3 AAG31117

Fri Aug 6 10:49:11 2004

CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2633 AA;

Query Match 100.0%; Score 2360; DB 4; Length 2633;
Best Local Similarity 100.0%; Pred. No. 1.9e-182; Indels 0; Gaps 0;
Matches 464; Conservative 0; Mismatches 0

QY 1 AEEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDRVPHGNETTKN 60
DB 2 AEEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDRVPHGNETTKN 61
QY 61 VVEEIAAPIIDSALIQYNGTIFAYGQTASGKTYTMGSEDHGLGVIPRAIHDFQIKKFP 120
DB 62 VVEEIAAPIIDSALIQYNGTIFAYGQTASGKTYTMGSEDHGLGVIPRAIHDFQIKKFP 121
QY 121 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVVYTSMAKWK 180
DB 122 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVVYTSMAKWK 181
QY 181 ITKGKSRHYGKTKMNRSSRSHTIFRMILESREKGEPSNCEGSKVSHLNLVDLAGSER 240
DB 182 ITKGKSRHYGKTKMNRSSRSHTIFRMILESREKGEPSNCEGSKVSHLNLVDLAGSER 241
QY 241 AAQTGAAGVRLKEGCNINRSLFILGVKIKLSDGVGFNDRVPHGNETTKN 300
DB 242 AAQTGAAGVRLKEGCNINRSLFILGVKIKLSDGVGFNDRVPHGNETTKN 301
QY 301 TRIICTITPVSPFDETLTALQFASAKYMKNTPYNEVSTDEALLKRYRKEIMDLKKOLEE 360
DB 302 TRIICTITPVSPFDETLTALQFASAKYMKNTPYNEVSTDEALLKRYRKEIMDLKKOLEE 361
QY 361 VSLTRAQAMEKDQALQLEEKDLQVNEKIEINLTMLVTSSSLTLQQLKAKRKERV 420
DB 362 VSLTRAQAMEKDQALQLEEKDLQVNEKIEINLTMLVTSSSLTLQQLKAKRKERV 421
QY 421 TWCLGKINKMKNNSYADQFNIPNTITTKTKLSINLREIDESV 464
DB 422 TWCLGKINKMKNNSYADQFNIPNTITTKTKLSINLREIDESV 465

RESULT 2
AAM39097
ID AAM39097 standard; protein; 2663 AA.
XX AAM39097;
AC AAM39097;
XX 22-OCT-2001 (first entry)
XX Human polypeptide SEQ ID NO 2242.
DE Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX Homo sapiens.
OS
XX WO200153312-A1.
PN
XX 26-JUL-2001.
PD

XX 26-DEC-2000; 2000WO-US034263.
XX 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR N-PSDB; AAI58253.
XX Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
XX Example 4; SEQ ID NO 2242; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nontropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX Sequence 2663 AA;

Query Match 100.0%; Score 2360; DB 4; Length 2663;
Best Local Similarity 100.0%; Pred. No. 1.9e-182; Indels 0; Gaps 0;
Matches 464; Conservative 0; Mismatches 0

QY 1 AEEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDRVPHGNETTKN 60
DB 2 AEEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDRVPHGNETTKN 61
QY 61 VVEEIAAPIIDSALIQYNGTIFAYGQTASGKTYTMGSEDHGLGVIPRAIHDFQIKKFP 120
DB 62 VVEEIAAPIIDSALIQYNGTIFAYGQTASGKTYTMGSEDHGLGVIPRAIHDFQIKKFP 121
QY 121 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVVYTSMAKWK 180
DB 122 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVVYTSMAKWK 181
QY 181 ITKGKSRHYGKTKMNRSSRSHTIFRMILESREKGEPSNCEGSKVSHLNLVDLAGSER 240
DB 182 ITKGKSRHYGKTKMNRSSRSHTIFRMILESREKGEPSNCEGSKVSHLNLVDLAGSER 241
QY 241 AAQTGAAGVRLKEGCNINRSLFILGVKIKLSDGVGFNDRVPHGNETTKN 300
DB 242 AAQTGAAGVRLKEGCNINRSLFILGVKIKLSDGVGFNDRVPHGNETTKN 301
QY 301 TRIICTITPVSPFDETLTALQFASAKYMKNTPYNEVSTDEALLKRYRKEIMDLKKOLEE 360
DB 302 TRIICTITPVSPFDETLTALQFASAKYMKNTPYNEVSTDEALLKRYRKEIMDLKKOLEE 361

QY 361 VSLTEAQAAMEKDQQLAEKDLQKQVQNEKIENLTRLMLVTSSSLTLOQELKAKRRV 420
 Db 362 VSLTEAQAAMEKDQQLAEKDLQKQVQNEKIENLTRLMLVTSSSLTLOQELKAKRRV 421
 QY 421 TWCLGKINKMKNKNYADQFNIPNTITTKTKLSINLLREIDESV 464
 Db 422 TWCLGKINKMKNKNYADQFNIPNTITTKTKLSINLLREIDESV 465

RESULT 3

AA040883
 ID AAM40883 standard; protein; 2688 AA.
 XX AAM40883;
 AC AAM40883;
 XX
 DT 22-OCT-2001 (first entry)
 DE Human polypeptide SEQ ID NO 5814.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US034263.
 XX
 PR 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-00523317.
 PR 20-JUN-2000; 2000US-00598042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX
 PA (HYSE-) HYSE INC.
 XX
 PI Tang YN, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren P, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
 PI Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR N-PSDB; AAI60039.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.
 XX
 PS Example 2; SEQ ID NO 5814; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
 CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral neuropathies, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification

XX
 SQ

Sequence 2688 AA;

Query Match 95.4%; Score 2251.5; DB 4; Length 2688;
 Best local similarity 96.6%; Pred. No. 1.3e-173;
 Matches 451; Conservative 2; Mismatches 11; Indels 3; Gaps 3;

QY 1 AEEGAVAVCVVRPLNSREESIGETAQVYWKTDNNVIYQVDGSKSFNDFVFGNETTKN 60
 Db 23 AEEGAVAVCVVRPLNSREESIGETAQVYWKTDNNVIYQVDGSKSFNDFVFGNETTKN 82
 QY 61 VVEETAAPIIDSALIOQYNGTIFAYGOTASGKTVMGSEDLGLVPIRA-IHDIF-OKIKK 118
 Db 83 VVEETAAPIIDSALIOQYNGTIFAYGOTASGKTVMGSEDLGLVPIQGFHGFHSOKIXE 142
 QY 119 -FPDREFLLRVSYMEIYNETITDLCGTQKMKPLIIRREDVNRNYYVADLTTEEYVYSEMA 177
 Db 143 VFLDREFLLRVSYMEIYNETITDLCGTQKMKPLIIRREDVNRNYYVADLTTEEYVYSEMA 202
 QY 178 LKWTGKESRHYGETKMKORSRSHITFRMILESEKGEPCNCEGSKVSHNLVLDLAG 237
 Db 203 LKWTGKESRHYGETKMKORSRSHITFRMILESEKGEPCNCEGSKVSHNLVLDLAG 262
 QY 238 SRAAQTGAAGVRLKEGNCNINRSFLFGQVVKLSGQVGGFVNYRDSKLTQLQNSLGG 297
 Db 263 SRAAQTGAAGVRLKEGNCNINRSFLFGQVVKLSGQVGGFVNYRDSKLTQLQNSLGG 322
 QY 298 NPKTRIICTITPVSFDETLTALQFASTAKYKNTPYNEVSTDEALLKRYRKEIMDLKKQ 357
 Db 323 NPKTRIICTITPVSFDETLTALQFASTAKYKNTPYNEVSTDEALLKRYRKEIMDLKKQ 382
 QY 358 LEEVSLTEAQAAMEKDQQLAEKDLQKQVQNEKIENLTRLMLVTSSSLTLOQELKAKRK 417
 Db 383 LEEVSLTEAQAAMEKDQQLAEKDLQKQVQNEKIENLTRLMLVTSSSLTLOQELKAKRK 442
 QY 418 RRVTWCLGKINKMKNKNYADQFNIPNTITTKTKLSINLLREIDESV 464
 Db 443 RRVTWCLGKINKMKNKNYADQFNIPNTITTKTKLSINLLREIDESV 489

RESULT 4

AA01632
 ID AAY01632 standard; protein; 2954 AA.
 XX
 AC AAY01632;
 XX
 DT 22-JUN-1999 (first entry)
 XX
 DE Amino acid sequence of centromere-associated protein-E (CENP-E).
 XX
 KW CENP-E; centromere-associated protein-E; ATPase activity;
 KW plus end-directed microtubule motor activity; chromosome congression;
 KW microtubule binding activity; chromosome movement; mitosis;
 KW cell proliferation; tumor; metastasis; vascular malfunction;
 KW inflammatory disease; immune disease; angiogenesis; hypertension;
 KW restenosis; fungal infection; selective herbicide; fungicide;
 KW insecticide; plant growth regulator; activator; cancer cell marker.
 XX
 OS Xenopus sp.
 XX
 PN WO9913061-A1.
 XX
 PD 18-MAR-1999.
 XX
 PF 10-SEP-1998; 98WO-US019231.
 XX
 PR 11-SEP-1997; 97US-0058645P.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Wood KW, Sakowicz R, Goldstein LSB, Cleveland DW;
 XX
 DR WPI; 1999-229233/19.

DR N-PSDB; AAX26819.
XX PT Centromere-associated protein-E and related nucleic acid.
XX PS
XX PS
XX Claim 5; Page 66-67; 77pp; English.
XX CC The present sequence represents CENP-E (centromere-associated protein-E) of
XX CC Xenopus. The protein has at least one of plus end-directed microtubule
XX CC motor activity, Arpase (adenosine triphosphatase) activity and
XX CC microtubule binding activity. CENP-E is the motor that powers chromosome
XX CC movement toward microtubule plus ends and is essential for congression of
XX CC chromosomes during mitosis. Modulators of CENP-E can thus control cell
XX CC proliferation. Agents that modulate CENP-E activity are lead therapeutic,
XX CC bioagricultural and diagnostic agents, e.g. for treatment of metastases;
XX CC cell proliferation (typical of many examples are tumors and metastases;
XX CC vascular malfunction; inflammatory and immune diseases; angiogenesis;
XX CC hypertension; restenosis; and fungal infections), also as plant-
XX CC protection agents (selective herbicides, fungicides and insecticides) and
XX CC plant growth regulators or activators for improving yields. CENP-E is
XX CC also a diagnostic marker for dividing cells, including cancer cells
XX CC
SQ Sequence 2954 AA;

Query Match 65.6%; Score 1549; DB 2; Length 2954;
Best Local Similarity 66.0%; Pred. No. 2.7e-116;
Matches 312; Conservative 63; Mismatches 86; Indels 12; Gaps 5;

QY 1 AEGAVAVCVVRPLNREBSLGTAYVYKTDNNVIYQVDSKSFNDFRVFHNGETTKN 60
DB 2 SEGDAVAVCVVRPLNREBSLGTAYVYKTDNNVIYQVDSKSFNDFRVFHNGETTKN 59
QY 61 VYEEIAPIIDSAIQVNGTIFAYGQTASGKTYTMWGSSEHLGVIPRAIHDFQKIKFP 120
DB 60 IYQEIAPVIRKSLQNGTIFAYGQTSSGKTYTMWGTSPNSLGIIPQAIQVFKIIEIP 119
QY 121 DREPLLRVSYMEIYNETITDLCGQTKMKPLIREDVNRNRYVADLTVEVYVTSMAIKW 180
DB 120 NREPLLRVSYMEIYNETITDLCGQTKMKPLIREDVNRNRYVADLTVEVYVTSMAIKW 179
QY 181 ITKGEKSHYGETKQNRSSRSHTIFRMILSRKGEPS---NCEGSKVSHLNLVDLAG 237
DB 180 IKGKGNRHYGETKQNRSSRSHTIFRMILSRKGEPS---NCEGSKVSHLNLVDLAG 239
QY 238 SERAAQTGAAGVRLKEGNCINRSFILGVQIKKLSGQGVGFYINRDSKLTILQNSLGG 297
DB 240 SERASQTGAAGVRLKEGNCINRSFILGVQIKKLSGQGVGFYINRDSKLTILQNSLGG 299
QY 298 NPKTRIICTITPVSFDETLTALQFASTAKYMKNTPYNEVSTDEALLKRYKEIMDLKKQ 357
DB 300 NAKTVIICTITPVSFDETLTALQFASTAKYMKNTPYNEVSTDEALLKRYKEIMDLKKQ 359
QY 358 LE--EVSLETRAQAMEKQQLAEKDLQKQVNEKLENLRLMTVSSSLTLOQELKAK 415
DB 360 LENLESSETQAQAMAKSEHTQLAEIKQLHREDEDRILWLTINIVASSQES-QDQVRK 418
QY 416 RKRRTVWGLKGNKMSNYADQFN---IPTNITTKTKLSINLLRIDEVS 464
DB 419 RKRRTVWGLKGNKMSNYADQFN---IPTNITTKTKLSINLLRIDEVS 471

RESULT 5
ABUS3125
ID ABUS3125 standard; protein; 366 AA.
XX AC ABUS3125;
XX AC
XX 15-APR-2003 (first entry)
XX DE Intracellular trafficking-associated DKFphtes3_26g22 homologue #2.
XX KW Human; gene therapy; vaccine; disease treatment; detection.
XX OS Homo sapiens.

XX PN WO200112659-A2.
XX PD 22-FEB-2001.
XX PF 18-AUG-2000; 2000WO-IB001496.
XX PR 18-AUG-1999; 99US-0149499P.
XX PR 28-SEP-1999; 99US-0156503P.
XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX PI Wiemann S;
XX WIPI; 2001-327840/34.
XX CC Nucleic acids having the sequences of clones isolated from libraries of
XX CC different human tissues, useful in recombinant DNA methodologies.
XX CC
PS Example III; Page 745; 1095pp; English.

XX CC This invention describes novel polynucleotides and polypeptides isolated
XX CC from human cDNA libraries which can be used for gene therapy or in
XX CC vaccines. The polynucleotides of the invention and antibodies encoded by
XX CC them may be used in the prevention, diagnosis and treatment of diseases
XX CC associated with inappropriate polypeptide expression. The products of the
XX CC invention may also be used to identify modulators of expression and
XX CC activity and to down regulate expression and activity. The antibodies of
XX CC the invention may also be used as diagnostic agents for detecting the
XX CC presence of polypeptides in samples. This sequence represents a homologue
XX CC of a polypeptide described in the disclosure of the invention

SQ Sequence 366 AA;
Query Match 34.7%; Score 820; DB 4; Length 366;
Best Local Similarity 50.5%; Pred. No. 3.4e-58;
Matches 187; Conservative 42; Mismatches 115; Indels 26; Gaps 7;

QY 11 RVRPLNREBSLGTAYVYKTDNNVIYQVDSKSFNDFRVFHNGETTKNRYVEE 64
DB 1 RCRPLNREBSLGTAYVYKTDNNVIYQVDSKSFNDFRVFHNGETTKNRYVEE 60
QY 65 IAAPIIDSAIQVNGTIFAYGQTASGKTYTMWGSSEHLGVIPRAIHDFQKIKFP 120
DB 61 VAHIVDDCFYNGTIFAYGQTSSGKTYTMWGTSPNSLGIIPQAIQVFKIIEIP 120
QY 121 DRE---FLLRVSYMEIYNETITDLCGQTKMKPLIREDVNRNRYVADLTVEVYVTSMA 176
DB 121 EKDHDFWHVKSVMYMEIYNETITDLCGQTKMKPLIREDVNRNRYVADLTVEVYVTSMA 180
QY 177 ALKWTKEKSHYGETKQNRSSRSHTIFRMILSRKGEPS---NCEGSKVSHLNLVDLAG 236
DB 181 ACHWITQGNRHYGETKQNRSSRSHTIFRMILSRKGEPS---NCEGSKVSHLNLVDLAG 236
QY 237 GSERAAQTGAAGVRLKEGNCINRSFILGVQIKKLSGQGVGFYINRDSKLTILQNSL 290
DB 237 GSERVNRFCAGQRLKEGNCINRSFILGVQIKKLSGQGVGFYINRDSKLTILQNSL 296
QY 291 LQNSLGGNPKTRIICTITPVSFDETLTALQFASTAKYMKNTPYNEVSTDEALLKRY 348
DB 297 LQDSLGGNCKTCMIACIWPADNRYVEETLSTLRADRAKNIKNKPOINDPCAMALWRRYH 356
QY 349 KEIMDLKKQL 358
DB 357 EQIDMKHQL 366

RESULT 6
ABUS3208
ID ABUS3208 standard; protein; 348 AA.
XX AC ABUS3208;
XX AC

DT 14-APR-2003 (first entry)
 DE Human cell cycle-associated DKFzptes3_35b4 homologue #7.
 DE Human; gene therapy; vaccine; disease treatment; detection.
 DE Homo sapiens.
 OS WO200112659-A2.
 PN 22-FEB-2001.
 PP 18-AUG-2000; 2000WO-IB001496.
 PR 18-AUG-1999; 99US-0149499P.
 PR 28-SEP-1999; 99US-0156503P.
 XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
 XX Wiemann S;
 PI WPI; 2001-327840/34.
 DR Nucleic acids having the sequences of clones isolated from libraries of
 PT different human tissues, useful in recombinant DNA methodologies.
 PS Example III; Page 829; 1095pp; English.
 XX This invention describes novel polynucleotides and polypeptides isolated
 CC from human cDNA libraries which can be used for gene therapy or in
 CC vaccines. The polynucleotides of the invention and antibodies encoded by
 CC them may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate polypeptide expression. The products of the
 CC invention may also be used to identify modulators of expression and
 CC activity and to down regulate expression and activity. The antibodies of
 CC the invention may also be used as diagnostic agents for detecting the
 CC presence of polypeptides in samples. This sequence represents a homologue
 CC of a polypeptide described in the disclosure of the invention
 XX Sequence 348 AA;
 SQ
 Query Match 33.1%; Score 781; DB 4; Length 348;
 Best Local Similarity 51.1%; Pred. No. 4.8e-55;
 Matches 178; Conservative 38; Mismatches 106; Indels 26; Gaps 7;
 QY 11 RVRPLNSREESLGETAQVYV-----KTDNNVIYQVDSKSFNEDRVFHNCTTKNYYEE 64
 DB 1 RCRPLNERINDGSCVQVQWPTGKTVHNGHGDSPKSFTEHVFWMNCTQSDVDYT 60
 QY 65 TAAPIDSALQYNGTIFAYGQTASGKTYTMG-----SEDLGLVTPRAIHDFQIKKFP 120
 DB 61 VAHPVDDCFHGYNCTIFAYGQTGSKTYTMGPGGHPDHMGIIIPRCHDIFDRDKFO 120
 QY 121 DRE---FLLRVSYMEIYNETITDLIC-GTQKKPLIIRDVNRNVIYVADLTBEVYVTSEM 176
 DB 121 EKDDHDFWVKCSMEIYNEEYDLCNPFQPMKPLNIHEHPNMGFYVQCTEFHVCSEYD 180
 QY 177 ALKWTITKESRHYGETKMNQSRSSHTIFRMILESRKEGSPNCEGSKYVSHLNLVLA 236
 DB 181 ACHWTQGNKRVHVAATNNDHSSSHTIFTIHVEQRHK-----QCDEHVCBSKMLVLA 236
 QY 237 GSERAQTAAGVRLKEGNCINRSFILQGVIKKLSGQV-----GGFINVRDSKLTRI 290
 DB 237 GSERVNRGTGAEQRLKEGNCINQSLTTLGNVINALADGQTKMYGCGHGIPIYRDSKLTL 296
 QY 291 LQNSLGGNPKRIITITPV--SFDETLTALQFASHTAKYMKNTPTVYNE 336
 DB 297 LQDSLGGNCKTMIACIWPADWNVEETLSTLRYADRAKNIKNKPQINE 344
 RESULT 7
 ABB62322
 ID ABB62322 standard; protein; 2013 AA.

XX ABB62322;
 XX 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 13758.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 DE pharmaceutical.
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 PP 27-SEP-2001.
 PR 23-MAR-2001; 2001WO-US009231.
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers BW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL06425.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 PS Disclosure; SEQ ID NO 13758; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 2013 AA;
 SQ
 Query Match 33.0%; Score 778.5; DB 4; Length 2013;
 Best Local Similarity 44.3%; Pred. No. 9.3e-54;
 Matches 190; Conservative 66; Mismatches 124; Indels 49; Gaps 12;
 QY 5 AVAVCVVRPLNSREESLGETAQVYVTKT-DNNVIYQVDP-SKSFNDRFVHGNCTTKNYY 62
 DB 8 SIQVCIKVRPCPEGLTSL-----WQVKERRSIHLADSHAEPYVDFVDFDEGASNQEVF 60
 QY 63 BEIAPIDSALQYNGTIFAYGQTASGKTYTMGSEDLGLVTPRAIHDFQIKKFPDR 122
 DB 61 DRMAKHIVHACQMGNGTIFAYGQTSSGKTYTMGDEQNGPVMVLAKEIFQOISSETER 120
 QY 123 EFLLRVSYMEIYNETITDLICGTQKKPLIIRDVNRNVIYVADLTBEVYVTSEM-ALKWI 181
 DB 121 DFLRVGVIEIYNEKIYDILL--NKNQDLKTHESNGIWNVN--CECIITSEVDLLRLIL 176
 QY 182 TGKSKSRHYGETKMNQSRSSHTIFRMILESRKEGSPNCEGSKYVSHLNLVLAGSERA 241
 DB 177 CLGNKERTVGTETNWNERSRSHAFIKIIIESR-KSDHSD-DDAVIQSVNLNVLVLAGSERA 234
 QY 242 AQTGAAGVRLKEGNCINRSFILQGVIKKLSGQVGGFINTRDSKLTRILQNSLGGNPKT 301
 DB 235 DQTGARGARLKEGGHINKSLFLSNVTKLSENADNRFTNYRDSKLTRILQASLGGNAFT 294
 QY 302 RIICITITPVSFDETLTALQFASHTAKYMKNTPTVYNEVSTDEALLKRYKEINDLKKQLEEV 361

Db 295 SIITIKPSIMESQSTSFATRAKKIRIKPQVNMVSDATMMKRLEREIKVLKDLAE- 353

QY 362 SLETRAQAMEKDQLAQLLEKDLLOKQVNEKIENLTRM-----LVTSSSLTLQOELKA 414

Db 354 -----EER-----KQENQKVHELRQIKHDMKHLICGHSLS-----DKG 388

QY 415 KRRKRRVTWC 423

Db 389 QQRKRRVTWC 397

RESULT 8

ABB65183

ID ABB65183 standard; protein; 677 AA.

XX

AC ABB65183;

XX

XX 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 22341.

XX

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX

OS Drosophila melanogaster.

XX

XX WO200171042-A2.

PN

XX

PD 27-SEP-2001.

XX

XX 23-MAR-2001; 2001WO-US009231.

XX

XX 23-MAR-2000; 2000US-0191637P.

PR

PR 11-JUL-2000; 2000US-00614150.

XX

XX (PBKE) PE CORP NY.

PA

XX

XX Venter JC, Adams M, Li PWD, Myers EW;

PI

XX

XX WPI; 2001-656860/75.

DR

DR N-PSDB; ABL09286.

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signaling and cell-cell

PT interactions.

XX

XX Disclosure; SEQ ID NO 22341; 2lpp + Sequence Listing; English.

PS

XX

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

CC ABB72072). The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 677 AA;

Query Match 30.4%; Score 716.5; DB 4; Length 677;

Best Local Similarity 40.8%; Pred. No. 2.2e-49;

Matches 189; Conservative 73; Mismatches 166; Indels 35; Gaps 13;

QY 6 VAVCVVRPLNRSRESLGETAQVYWKTDNNVIYQV-----DGSKSFNFDFVHGNETT 58

Db 21 VRVWVTRPMDKNLSAGALSALSVDKINRAITVMKPNATANEPPKTYTFDNDVGGSNQ 80

QY 59 KNYVEIAAPIIDLSALQGVNGTIFAYGQTASGKTYTMMGSEDH---LGVIPRAIHDIFQK 115

Db 81 MDLYVDTARPIVDKVLGVNGTILAYGQTGKTYTMSGNPDSPQTKIIPNAFAHIFGH 140

QY 116 IKKFPD-REFLLRVSYMEIYNETITDLCQTKMKPLIIRVDVNRNRYVADLTVEVYVTS 174

Db 141 IAKAKENQFLVRSVSYMEIYNEVRDLL-GKDVGKSLVKERPDIGVFVKDLGGYVYVNA 199

QY 175 EMALKWITTKGKSRHYGETKQNRSSRSHTIFRMILESRKKGPSNCEGSKVYSHLNLVD 234

Db 200 DDLNIMRLGNKRAVGATKQNCQSSRSASHAIFSTITVERSELGEGD--VQVHRMGKQLQVLD 257

QY 235 LAGSERAQAOTGAAGVRLKEGCNINRSILFGLQVVKKLSGQVGVGFNRYRDSKLTIRLONS 294

Db 258 LAGSERQSKTQASGQRLKEATKINLSLVGNVISALVDGK-STHPIYRNSKLTIRLQDS 316

QY 295 LGGNPKTRIICTTPV--SFDETILTAQFASTAKYMKNTPYVNEVSTDEALLKRYRKEIM 352

Db 317 LGGNSKTVMCATISPADSNYMETISITLRYASRAKNIQNRHINEEPKD-ALLRHFOSEIA 375

QY 353 DLKKQLEE-VSLETRAQAMEKDQLAQLLEKDLLOKQVNEKIENLTRLMTVSSSLTLQO- 410

Db 376 RLKQLEEGDLSLEEPSSE-----EBEDTADDELEAPLE-----IELESSTIQAV 421

QY 411 ELKAKRRTVTC-LGKINKMKNSYADQFNIPNTITTKTKL 452

Db 422 EKPKKKREKTDAAKELAKRKNHOKHEIHAKEQETLRNKL 464

RESULT 9

AAE17786

ID AAE17786 standard; protein; 1029 AA.

XX

AC AAE17786;

XX

XX 07-MAY-2002 (first entry)

XX

DE Human kinesin superfamily motor protein, HsKif17.

XX

XX Human; kinesin superfamily motor protein; HsKif17; autoimmune disease;

KW cellular proliferation; therapy; cancer; hyperplasia; graft rejection;

KW cardiac hypertrophy; immune disorder; inflammation; immunosuppressive;

KW angioplasty; arthritis; restenosis; inflammatory bowel disease;

KW cytostatic; vasotropic; antiarthritic; antiinflammatory.

XX

OS Homo sapiens.

XX

XX WO200198314-A2.

PN

XX

PD 27-DEC-2001.

XX

XX 20-JUN-2001; 2001WO-US019811.

XX

XX 20-JUN-2000; 2000US-00597602.

PR

XX (CYTO-) CYTOKINETICS INC.

XX

XX Beraud C, Freedman R;

PI

XX

XX WPI; 2002-147789/19.

DR

DR N-PSDB; AAD28568.

XX

XX Novel human kinesin superfamily motor protein, HsKif17, useful for

PT identifying modulators that may be used to treat cell proliferative

PT disorders, e.g., cancer, hyperplasias, restenosis, autoimmune disease,

PT arthritis, graft rejection.

XX

PS Claim 11; Fig 2; 64pp; English.

XX

CC The present invention relates to an isolated novel human kinesin

CC superfamily motor protein, HsKif17. HsKif17 is useful for screening

CC modulators of HsKif17, comprising contacting biologically active HsKif17

CC with a candidate agent in a test and control concentration and assaying

CC for the level of HsKif17 activity where a change in activity between the

CC test and control concentration indicates a modulator, and where screening

CC occurs in a multiwell plate as a part of high-throughput screen. The

CC modulators identified are useful in treatment of cellular proliferation

Matches 185; Conservative 75; Mismatches 165; Indels 68; Gaps 15;

QY 5 AVAVCVVRPLNSREESLG-ETAQVYWKTD-----NNVIYQVDSKGFNDFRVHGN 56

Db 5 AVKVVRCRPMNQREEL--RCQVTVVDCARACQIQNPGADEPKQFTFDGAYVDH 62

QY 57 TTKNNVYEIAAPLIDSAIQNGYNTIFAYGQTASGKTYTMMGSD--HLGVIPRAIHDF 113

Db 63 VTBEQIYNEIAYPLVEGTYGNTIFAYGQTGSGKSTFMQGLPDPSPQSGIIPRAEFHF 122

QY 114 QKIKKPPDFRFLRVSYMEIYNETITDLCGTOKMKPLIREDVNRNYYADLTTEEVTY 173

Db 123 ESVQCAENTKFLVRSYLEIYNEIDVDLLGADTKQK-LELKEHPEKGVYVKGUSMTVHS 181

QY 174 SEMALKWITKGEKSRHYGETKMNQSRSHITFRMILESR---EKGEPSNCEGSKVYSHL 230

Db 182 VAQCEHIMETGKNRSVGYTLNMKDSRSHISITSIEMSAVDERG-----KDLHRAKGL 236

QY 231 NLVDLAGSRAAQTGAAGVRLKEGCNINRSLFILGVQVKKLSGQVGGFINYRDSKLTRI 290

Db 237 NLVDLAGSRSQKGTGATGERLKEATKINLSLSALGNVISALVDGRC-KHVPYRDSKLTRL 295

QY 291 LQNSLGNPKTRITCITPV--SPDETLTALQFASAKYMKNTPYNEVSTDEALLKRYR 348

Db 296 LQDSLGGTKTLVACLSPADNNYDETLSTLRANRKNIRNRPINEDPKD-ALLREYQ 354

QY 349 KEIMDLKKOLEVSLERAQAMEKDQALQL-----BEKDLQKV-----388

Db 355 ERIKKLKAII-----TQMSPSLSALLSRQVDPVQVEEKLPPQVIQDHMEAEK 406

QY 389 ---ONEKIENLRLMTVSSLTLOQLKAKRKRRTVWCLGKINMKNSYVADQFNPTNI 445

Db 407 QLIREEVEERLARU---KADYKAEQESRARLEEDIT-----AMRNSYDVRSLTEENL 456

QY 446 TTKTHK-LSINLL 457

Db 457 RKETEAVLQGVGL 469

RESULT 11

ID ADD49938 standard; protein; 1232 AA.

XX AC ADD49938;

XX DT 15-JAN-2004 (first entry)

XX DE Human lung specific tumour antigen L 1447p.

XX KW Human; lung cancer antigen; cytostatic; lung cancer; gene therapy; vaccine; T-cell; tumour.

XX OS Homo sapiens.

XX PN US2003194764-A1.

XX 16-OCT-2003.

XX PF 04-APR-2002; 2002US-00116712.

XX PR 05-APR-2001; 2001US-0282289P.

XX PR 05-OCT-2001; 2001US-0327511P.

XX PA (CORI-) CORIXA CORP.

XX PI Bangor CS, Switzer A;

XX DR WPI; 2003-844452/78.

XX DR N-PSDB; ADD49936.

XX PT New isolated polypeptides and polynucleotides useful for diagnosing, preventing and treating cancer, particularly lung cancer.

XX

PS Claim 1; SEQ ID NO 670; 250pp; English.

XX The invention relates to an isolated polynucleotide (a) comprising any of the 666 fully defined nucleotide sequences appearing as ADD49269 - ADD49935, ADD49936 and ADD49938, complements of (a); sequences of at least 20 contiguous residues of (a); sequences that hybridise to (a) under highly stringent conditions; sequences having at least 75 or 90% identity to (a); or degenerate variants of (a). Also included are an isolated polypeptide (b) (comprising: sequences encoded by the new polynucleotide; any of the 4 amino acid sequences fully defined in the specification; or sequences having at least 70 or 90% identity to the sequence in (a) or (b)), an expression vector comprising the above polynucleotide operably linked to an expression control sequence, a host cell transformed or transfected with the above expression vector, an isolated antibody, or its antigen-binding fragment, that specifically binds to the above polypeptide, an oligonucleotide that hybridises to the above-mentioned nucleotide sequences under highly stringent conditions, a fusion protein comprising at least one polypeptide cited above, detecting the presence of a cancer in a patient (comprising: obtaining a biological sample from the patient; contacting the biological sample with a binding agent that binds to the polypeptide, or with the oligonucleotide cited above; detecting in the sample an amount of the polypeptide that binds to the binding agent, or an amount of a polynucleotide that hybridises to the oligonucleotide; and comparing the amount of polypeptide, or polynucleotide that hybridises to the oligonucleotide, to a predetermined cut-off value and then determining the presence of a cancer in the patient), a method for stimulating and/or expanding T-cells specific for a tumour protein (comprising contacting T-cells with the above polypeptide, polynucleotide or antigen-presenting cells that express the polynucleotide, under conditions and for a time sufficient to permit the stimulation and/or expansion of T-cells), an isolated T-cell population comprising T-cells prepared by the method, a composition comprising a first component selected from physiological carriers and immunostimulants, and a second component selected from the above polypeptide, polynucleotide, antibody, fusion protein, T-cell population and antigen-presenting cells that express the above polypeptide, stimulating an immune response in a patient (comprising administering to the patient the above composition) treating lung cancer in a patient (comprising administering to the patient the above composition and a diagnostic kit (comprising: at least one oligonucleotide cited above; or at least one antibody cited above and a detection reagent, where the detection reagent comprises a reporter group). The composition and methods are useful in diagnosing, preventing and treating cancer, particularly lung cancer. The present sequence is a lung cancer-associated antigen of the invention.

Sequence 1232 AA;

Query Match 29.6%; Score 697.5; DB 7; Length 1232;

Best Local Similarity 36.2%; Pred. No. 1.8e-47;

Matches 179; Conservative 79; Mismatches 178; Indels 59; Gaps 11;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKGFNDFRVHGNETTKNYEE 64

Db 10 VRVALRCRPLVPKREISEGCGMCLSFVPEGPVVGTD--KSFTYDFVDFPSTEQEEVNT 67

QY 65 IAAPIIDSAIQNGYNTIFAYGQTASGKTYTMMG-----SEHGLGVIPRAIHDFQKIK 117

Db 68 AVAPLKGVEKYNATVLAYGQTGSGKTYSGMGAYTAQENETVGVIPVQLLFKEID 127

QY 118 KFPDPREFLLRVSYMEIYNETITDLCGTOKMKPLIREDVNRNYYADLTTEEVTYSEMA 177

Db 128 KKSDFEFTLVKSYLEIYNEIIDLDCPSREKAKINREDPKGKIVGLTEKTVLVALDT 187

QY 178 LKMITKGEKSRHYGETKMNQSRSHITFRMILESRKEGPPSNCEGSKVYSHLNLVDIAG 237

Db 188 VSCLEQGNNSRTVASTAMNSQSRSHAFITISLEQRKSD-----KNSFSRSLHLVDIAG 243

QY 238 SERAAQTGAAGVRLKEGCNINRSLFILGVQVKKLSGQVGGFINYRDSKLTTRILQNSLGG 297

Db 244 SEROKTKAEGDRLKEGININRGLLCGNVISALGDDKGGFVPRDSKLTTRILQDSLGG 303

QY 298 NPKTRITCITPV--SPDETLTALQFASAKYMKNTPYNEVSTDEALLKRYRKEIMDLK 355

Db 304 NSHTMTACVSPADSNLEETLTLRYADRAKKNKPIVN-----IDPQTAEHLNHLK 355
 QY 356 KQLEEVSL-----ETRAQAMEKDQLAQLEEKDQLQVQNEKIENL 396
 Db 356 QVQVQLVQLLQAHGGTLPQSITVPESENLSLMKQ--SLVEENEKLSRGSEAAAGOT 413
 QY 397 TRML-----VTSSITLQQLKAKRRVTWCLGKINK-----MKNSYADQFNIPNTTKT 449
 Db 414 AQMLERIIITQEQANEMKNAKLEELRQHAACKLDLQKLIVETLEQELKENVEIICNQLLI 473
 QY 450 HKLSINLLREIDESV 464
 Db 474 TQLS-----DETV 481

RESULT 12

ADD18924

ID ADD18924 standard; protein; 1232 AA.

XX AC ADD18924;

XX DT 15-JAN-2004 (first entry)

XX DE Human disease related protein SeqID413.

XX KW human; disease state; cytostatic; antiinflammatory; ophthalmological;
 KW antiarteriosclerotic; vulnary; gene therapy;
 KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;
 KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;
 KW glucose transportation; catecholamine synthesis; iron transport;
 KW nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;
 KW retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;
 KW inflammatory condition; wound healing.

XX OS Homo sapiens.

XX PN WO2003018621-A2.

XX PD 06-MAR-2003.

XX PF 23-AUG-2002; 2002WO-GB003892.

XX PR 23-AUG-2001; 2001GB-00020558.

XX PR 05-OCT-2001; 2001GB-00024037.

XX PA (OXFO-) OXFORD BIOMEDICA UK LTD.

XX PI Kingsman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;

XX DR WPI; 2003-290046/28.

XX DR N-PSDB; ADD18925.

XX PT New substantially purified polypeptide, useful for diagnosing or treating
 PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion
 PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or
 PT wound healing.

XX PS Claim 25; SEQ ID NO 413; 424pp; English.

XX CC This invention relates to novel human genes and gene product which are
 CC implicated in certain disease states. Compounds which modulate the
 CC proteins of the invention may have cytostatic, antiinflammatory,
 CC ophthalmological, antiarteriosclerotic or vulnerary activities. The
 CC sequences of the invention may be useful for gene therapy. The invention
 CC may be useful for diagnosing or treating a hypoxia-regulated condition,
 CC such as tumorigenesis, angiogenesis, apoptosis, inflammation,
 CC erythropoiesis, or the biological response to hypoxia conditions
 CC including processes such as glycolysis, gluconeogenesis, glucose
 CC transportation, catecholamine synthesis, iron transport or nitric oxide
 CC synthesis. The disease includes cancer, ischaemic conditions, reperfusion
 CC injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,
 CC inflammatory conditions or wound healing. The present sequence is that of

CC a disease related protein of the invention.

XX SQ Sequence 1232 AA;
 Query Match 29.6%; Score 697.5; DB 7; Length 1232;
 Best Local Similarity 36.2%; Pred. No. 1.8e-47;
 Matches 179; Conservative 79; Mismatches 178; Indels 59; Gaps 11;

QY 6 VAVCVRVRLNSREESLG-ETAVYWKTDNNVYIQVDGSKSPNDRVFGHNKTNKYYEE 64
 Db 10 VRVALRCRLVPKPEISEGQCMCLSFVPGEPQVVGTD--KSFTYDFVDFDSTQEVEFNT 67
 QY 65 IAAPIIDSAIQGYNGTIFAYGQTASCKTYTMG-----SEHGLGVIPRAIHDFQKIK 117
 Db 68 AVAPLKGKGVKGNVAVLAYGQTGSKTYSMGAYTABOENETVGVIPRVIQLLFEKID 127
 QY 118 KFPDREFLLRVSYMEIYNETITDLLCGTQKMKPLIREDVNRNVYVADLTTEEVVYTSMA 177
 Db 128 KKSDPEFTLKVSYLEIYNEEILDLLCPSREKAIQINREDPKGKIKIVGLTEKTVLVALDT 187
 QY 178 LKWITKGEKSRHYGETKMQRSRSHITPRMILESEKGEPSNCEGSKVSHLNVLVDLAG 237
 Db 188 VSCLEQGNNSRTVASTAMNSQSRSRSHAITISLEQRKKSD----KNSSFRSKLHLVDLAG 243
 QY 238 SERAQTGAAGVRLKEGONINSLFILGOVKKLSGQVGGFINVEDSKLTRILQNSLGG 297
 Db 244 SERQKTKAEGDRLKEGININGLCLGNVISALGDDKKGGFVYRDSKLTRILQNSLGG 303
 QY 298 NPKTRIICITPV--SFDETLTALOPASTAKYMKNTPYNVNEVSTDEALLKRYRKEIMDLK 355
 Db 304 NSHTLMIAVCVSPADSNLEETLTLRYADRAKKNKPIVN-----IDPQTAEHLNHLK 355
 QY 356 KQLEEVSL-----ETRAQAMEKDQLAQLEEKDQLQVQNEKIENL 396
 Db 356 QVQVQLVQLLQAHGGTLPQSITVPESENLSLMKQ--SLVEENEKLSRGSEAAAGOT 413
 QY 397 TRML-----VTSSITLQQLKAKRRVTWCLGKINK-----MKNSYADQFNIPNTTKT 449
 Db 414 AQMLERIIITQEQANEMKNAKLEELRQHAACKLDLQKLIVETLEQELKENVEIICNQLLI 473

RESULT 13

ABG70992

ID ABG70992 standard; protein; 473 AA.

XX AC ABG70992;

XX DT 10-DEC-2002 (first entry)

XX DE Human target protein.

XX KW Microtubule stimulated ATPase; cellular proliferation; cancer; AIDS;
 KW hyperplasia; restenosis; cardiac hypertrophy; immune disorder; HIV;
 KW inflammation; autoimmune disease; arthritis; graft rejection; psoriasis;
 KW inflammatory bowel disease; proliferation; medical procedure; surgery;
 KW human immunodeficiency virus; acquired immunodeficiency syndrome;
 KW angioplasty; human; Hskif; kinesin family.

XX OS Homo sapiens.

XX PN US6440684-B1.

XX PD 27-AUG-2002.

XX PF 12-JUN-2000; 2000US-00592054.

XX PR 12-JUN-2000; 2000US-00592054.

XX PA (CYTO-) CYTOKINETICS INC.

XX PI Beraud C, Finer JT, Sakowicz R, Wood KW;
XX WPI; 2002-711529/77.
XX N-PSDB; ABS55162.
XX Screening for modulators of target protein having microtubule stimulated
PT ATPase activity e.g. kinesin family of protein, useful for treating
PT cancer, psoriasis, arthritis, human immunodeficiency virus (HIV)
PT infection.
XX Claim 3; Fig 6; 34pp; English.
XX The present invention relates to a new method of screening modulators of
CC target protein with microtubule stimulated ATPase activity. The method
CC involves contacting the target protein with an agent at 1st and 2nd
CC concentrations and determining the level of activity (e.g. binding or
CC ATPase activity) of target protein, where a difference between levels of
CC activity of target protein contacted with 1st and 2nd concentrations of
CC an agent indicates that an agent modulates activity of target protein.
CC The invention can be used for screening for modulators of target protein
CC having microtubule stimulated ATPase activity. The compounds identified
CC by method of the invention are useful for treating cellular proliferation
CC including cancer, hyperplasias, restenosis, cardiac hypertrophy, immune
CC disorders and inflammation. The compounds identified by the method are
CC also useful for treating autoimmune disease, arthritis, graft rejection,
CC inflammatory bowel disease, proliferation induced by medical procedures,
CC e.g. surgery, angioplasty etc. The compounds are also useful for treating
CC psoriasis. The compounds are useful for inhibiting human immunodeficiency
CC virus (HIV) and thus treating acquired immunodeficiency syndrome (AIDS).
CC The present amino acid sequence represents the human target protein of
XX the invention
XX Sequence 473 AA;
XX Query Match 29.4%; Score 695; DB 5; Length 473;
XX Best Local Similarity 39.7%; Pred. NO. 7.6e-48;
XX Matches 172; Conservative 70; Mismatches 155; Indels 36; Gaps 10;
QY 6 VAVCVVRPLNSRESIG-ETAYQVYKTDNNVIVQDGSKSFNEDRVPHGNETTKNVYEE 64
Db 8 VRVALRCRPLVPEKEISECQMLCFVPGEPQVVGWTD--KSFYDFVDPSTQEVEVNT 65
QY 65 IAAPIIDSAIGYNTIFAYGQTASGKTYTWMG-----GSDHLGVIPRAIHDFQKIK 117
Db 66 AVAPLIKGVFGYNATVLAAGTGTSGKTYSMGAYTAQEENPTVGVIPRVQLLKEID 125
QY 118 KFPDRFELLRVSYMEIYNETITDLLCGTKMKPLIREDVNRNVVADLTREVVYVTEMA 177
Db 126 KKSDFEFTLVKSYLBIYNEEILDLCPGREKAQINIREDPKEGIVGITEKTVLVALDT 185
QY 178 LKWITGKSRHYGTQKQNSRSHHTIFRMLESREKGEPCSGVKVSHNLNLDVLAG 237
Db 186 VSCLEQGNNSRTVASTAMNSQSSRSHAFITISLEQRKXSD-----KNSFSRKLHLVDLAG 241
QY 238 SRAAQTGAAGVRLKEGGINNSLSFILGVIVKLSGQGVGFYNYRDSKLTILQNSLGG 297
Db 242 SERQKKTGAEGRLKEGINNRLCLGNVISAIGDDKGGFVFPYRDSKLTLLQDSLGG 301
QY 298 NPKTRIICTITPV--SPDETILALQFAPSTAKYMKNTPPYVNEVSTDEALLKRYRKEIMDK 355
Db 302 NSHTLMIAVCSPADSNLEETLNLRYADRAKIKNKPVN-IDPQTAEHLNHLKQVY---- 356
QY 356 KQLEVSLETRAQAM-----EKDQLAQLEBKDLLQKQVNEKLENLTMLVTSSTLTQ 409
Db 357 QQLQVLLQAHGQTGPSITVSPENLQSLMEKNQSLVENEK---LSRGLSEAAQQT--- 411
QY 410 QELKAKRKRRTVW 422
Db 412 ----AQMLERLIW 420
RESULT 14

ABG70991
ID ABG70991 standard; protein; 522 AA.
XX
AC ABG70991;
XX
DT 10-DEC-2002 (first entry)
XX
XX Human Hskif4 construct protein.
XX
XX Microtubule stimulated ATPase; cellular proliferation; cancer; AIDS;
XX hyperplasia; restenosis; cardiac hypertrophy; immune disorder; HIV;
XX inflammation; autoimmune disease; arthritis; graft rejection; psoriasis;
XX inflammatory bowel disease; proliferation; medical procedure; surgery;
XX human immunodeficiency virus; acquired immunodeficiency syndrome;
XX angioplasty; human; Hskif4; kinesin family.
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH Region 1..22 /note= "N-terminal T7 epitope"
FT Region 496..516
FT Region /note= "C-terminal myc epitope"
FT Region 517..522
FT /note= "6-histidine residues at C-terminus"
XX US6440684-B1.
XX 27-AUG-2002.
XX 12-JUN-2000; 2000US-00592054.
XX 12-JUN-2000; 2000US-00592054.
XX (CYTO-) CYTOKINETICS INC.
XX Beraud C, Finer JT, Sakowicz R, Wood KW;
XX WPI; 2002-711529/77.
XX N-PSDB; ABS55161.
XX Screening for modulators of target protein having microtubule stimulated
PT ATPase activity e.g. kinesin family of protein, useful for treating
PT cancer, psoriasis, arthritis, human immunodeficiency virus (HIV)
PT infection.
XX Claim 2; Fig 4; 34pp; English.
XX The present invention relates to a new method of screening modulators of
CC target protein with microtubule stimulated ATPase activity. The method
CC involves contacting the target protein with an agent at 1st and 2nd
CC concentrations and determining the level of activity (e.g. binding or
CC ATPase activity) of target protein, where a difference between levels of
CC activity of target protein contacted with 1st and 2nd concentrations of
CC an agent indicates that an agent modulates activity of target protein.
CC The invention can be used for screening for modulators of target protein
CC having microtubule stimulated ATPase activity. The compounds identified
CC by method of the invention are useful for treating cellular proliferation
CC including cancer, hyperplasias, restenosis, cardiac hypertrophy, immune
CC disorders and inflammation. The compounds identified by the method are
CC also useful for treating autoimmune disease, arthritis, graft rejection,
CC inflammatory bowel disease, proliferation induced by medical procedures,
CC e.g. surgery, angioplasty etc. The compounds are also useful for treating
CC psoriasis. The compounds are useful for inhibiting human immunodeficiency
CC virus (HIV) and thus treating acquired immunodeficiency syndrome (AIDS).
CC The present amino acid sequence represents the human Hskif4 (kinesin
CC family) construct protein of the invention
XX Sequence 522 AA;
XX Query Match 29.4%; Score 695; DB 5; Length 522;
XX Best Local Similarity 39.7%; Pred. NO. 8.7e-48;
XX Matches 172; Conservative 70; Mismatches 155; Indels 36; Gaps 10;

QY 6 VAVCVVRPLNSRSLG-ETAQVYKTKDNNVIYQVDSKSFNFDRVFGHNETTKNYEE 64
 Db 30 VRVALCRPLVPKEISGQCMCLSFVPGEPQVVGTD--KSFTYDFVDFDSTQEVEFNT 87
 QY 65 IAPIIDSAIQNGNITIFAYGQTASGKTYTMMG-----SEHGLGVIPRAIHDFOKIK 117
 Db 88 AVAPLKGKFGYNATVAYGQTGSGKTYGSGGAYTAEGNEPTGVIPRVIQLLPKEID 147
 QY 118 KFPDRFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRVYVADLTVEEYVYVYSEMA 177
 Db 148 KKSDFEFTLKVSLEYNEIEILLCPREKAQINIREDPKEGKIVGLTEKTVLVALDT 207
 QY 178 LKWITGKSRHYGETKMNORSRSHITFRMILESRKGEPSCEGSKVSHLNLDVLAG 237
 Db 208 VSCLEQGNNSRTVASTAMNSQSSKSHAFITISLEQRKSD----KNSSFRSKLHLVDLAG 263
 QY 238 SERAOTGAAGVRLKGCNINRSLFILGOVKKLSGQVGGFYNRDSKLTIRLQNSLGG 297
 Db 264 SERQKTKRAGDRLKEGININRGLLCLGNVISALGDDKGGFVYRDSKLTIRLQNSLGG 323
 QY 298 NPKTRIICITIPV--SFDETLTALQFASHTAKYMKNTPYNEVSTDEALLKRYKEIMDLK 355
 Db 324 NSHTLMIACVSPADSNLEETLNTLYADRAKTKNPIVN-IDPQAEHLNHLKQVQV---- 378
 QY 356 KOLEEVSLETRAQAM-----EKDQLAQLLEKDLQKQVNEKIENLTRMLVTSSSLTLQ 409
 Db 379 QQLQVLLQAHGGTLPGSITVPSLENQSLMEKQSLVEENEK---LSRGLSEAAQOT--- 433
 QY 410 QELKAKRKERTVW 422
 Db 434 ----AQMLERIIV 442

RESULT 15

AAW48337
 ID AAM48337 standard; protein; 1038 AA.
 AC AAM48337;
 XX
 DT 18-APR-2002 (first entry)
 XX
 DE Human kinesin superfamily motor protein, KIF17.
 XX
 KW Human; kinesin superfamily; motor protein; KIF17; microtubule binding;
 KW nerve disease.
 XX
 OS Homo sapiens.
 XX
 PN JP2001333775-A.
 XX
 PD 04-DEC-2001.
 XX
 PF 24-MAY-2000; 2000JP-00153664.
 XX
 PR 24-MAY-2000; 2000JP-00153664.
 XX
 PA (SENT-) SENTAN KAGAKU GIJUTSU INCUBATION CENT KK.
 PA (UITY) UNIV TOKYO.
 XX
 DR WPI; 2002-135941/18.
 DR N-PSDB; ABA98475, ABA98176.
 XX
 PT A kinesin superfamily motor protein KIF17, useful for the prevention
 PT and/or treatment of nerve diseases caused by abnormal nerve cells.
 XX
 PS Claim 1; Page 12-15; 23pp; Japanese.
 XX
 CC The present sequence is human kinesin superfamily motor protein, KIF17.
 CC KIF17 is a microtubule binding motor protein which can be used for the
 CC prevention and/or treatment of nerve diseases caused by abnormal nerve
 CC cells
 XX

SQ Sequence 1038 AA;

Query Match 29.4%; Score 695; DB 5; Length 1038;
 Best Local Similarity 40.3%; Pred. No. 2.3e-47;
 Matches 172; Conservative 74; Mismatches 143; Indels 38; Gaps 12;
 QY 5 AVAVCVVRPLNSRSLG-ETAQVYKTKDNNVIYQVDSG-----KSFNDRVFGHNETT 58
 Db 5 SVKVVRCFPMKRRRELSCQSVTVVDSARGQCQFIONPGADEPPKQFTFDGAYIEHFT 64
 QY 59 KNVYERIAAPIIDSAIQNGNITIFAYGQTASGKTYTMMGSED---HLGVIPRAIHDFOK 115
 Db 65 EQIYNEIAVPLVEGVTEGYNGITIFAYGQTGSGKSFQMQLPDPCCQGIIPRAFEHVFS 124
 QY 116 IKKFPDRFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRVYVADLTVEEYVYVYSE 175
 Db 125 VQCAENTKFLVRASYLEIYNEIDVHDL--GADTKQRLKEHPEKGVYVVKLSMHTVHVA 183
 QY 176 MALKWITGKSRHYGETKMNORSRSHITFRMILE--SREKGEPSNCEGSKVSHLN 232
 Db 184 QCEVMEVETGKRAVGYTLMNKDSSRSHSIFINIEIYAVDERG-----KHLRAQKLN 238
 QY 233 VDLASERAAQTGAAGVRLKGCNINRSLFILGOVKKLSGQVGGFYNRDSKLTIRLQ 292
 Db 239 VDLASERQSKTGATGERLKEATKINLSLALGNVISALVDGRC-KHIFYRDSKLTIRLQ 297
 QY 293 NSLGGNPKTRIICITIPV--SFDETLTALQFASHTAKYMKNTPYNEVSTDEALLKRYKE 350
 Db 298 DSLGGNTKTLMAVCLSPADNNYDETSLTRYANRAKNIKKNPRINEDPKD-ALLREYQEE 356
 QY 351 IMDLKKQLEEVSLERAQAMEKDQLAQLLEKDLQKQVNEKIENLTRMLVTSSSLTLQ 410
 Db 357 IKELKAIL-----AQQMGFNLSALLSTQTPPGFVQSEE-----KLL----SPTTVQ 400
 QY 411 ELKAKRK 417
 Db 401 DTEAEKQ 407

Search completed: July 29, 2004, 09:35:14
 Job time : 28.9291 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:26:01 ; Search time 29.3136 Seconds
(without alignments)
4694.096 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_488

Perfect score: 2483

Sequence: 1 AEEGAVAVCVVRPLNSREE.....SDVFSNTLDTLSEIENPAT 487

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2483	100.0	2633	4	ABG06505	Abg06505 Novel hum
2	2483	100.0	2663	4	AAM39097	Aam39097 Human pol
3	2374.5	95.6	2688	4	AAM40883	Aam40883 Human pol
4	1584	63.8	2954	2	AAV01632	Aav01632 Amino aci
5	820	33.0	366	4	ABU53125	Abu53125 Intracell
6	781	31.5	348	4	ABU53208	Abu53208 Human cel
7	780.5	31.4	2013	4	ABBE6322	Abbe6322 Drosophil
8	716.5	28.9	677	4	ABB65183	Abb65183 Drosophil
9	713.5	28.7	1232	7	ADD49938	Add49938 Human lun
10	713.5	28.7	1232	7	ADD18924	Add18924 Human dis
11	710.5	28.6	1029	5	AAE17786	Aae17786 Human kin
12	710.5	28.6	1029	7	ADC10190	Adc10190 Human NOV
13	705.5	28.4	1234	5	ABG70993	Abg70993 Human Hsk
14	705	28.4	1034	3	AAG31112	Aag31112 Arabidops
15	705	28.4	1069	3	AAG31111	Aag31111 Arabidops
16	705	28.4	1121	3	AAG31110	Aag31110 Arabidops
17	703	28.3	1048	4	ABB59245	Abb59245 Drosophil
18	703	28.3	1066	6	AAG67418	Aag67418 Amino aci
19	703	28.3	1066	6	ABG72693	Abg72693 Fruitfly
20	702.5	28.3	1038	5	AAM48337	Aam48337 Human kin
21	702.5	28.3	1232	7	ADD49937	Add49937 Human kin
22	702.5	28.3	1232	7	ADD49932	Add49932 Human lun
23	702	28.3	522	5	ABG70991	Abg70991 Human Hsk
24	696.5	28.1	1232	5	ABG70990	Abg70990 Human Hsk
25	695	28.0	473	5	ABG70992	Abg70992 Human tar

ALIGNMENTS

RESULT 1

ABG06505
ID ABG06505 standard; protein; 2633 AA.

XX AC ABG06505;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #6496.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS70692.

XX New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity.

XX Claim 20; SEQ ID NO 36864; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
sequences. (I) is useful as hybridisation probes, polymerase chain
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
and in recombinant production of (II). The polynucleotides are also used
in diagnostics as expressed sequence tags for identifying genes.
XX genes. (I) is useful in gene therapy techniques to restore normal
activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging

Adb67093 Kinesin h
Aaw72746 Drosophil
Abb63485 Drosophil
Adb67088 Kinesin h
Aaul9569 Human dia
Abp51294 Human MDD
Aag31117 Arabidops
Aag31116 Arabidops
Ade5349 Rat Prote
Aae14400 Human kin
Aau79590 Human kin
Abr48222 Human bla
Adb80468 Ovarian c
Adc35116 Human bre
Aam78880 Human pro
Adb67089 Kinesin h
Adb67091 Kinesin h
Aaw72745 Drosophil
Aaw72744 Drosophil
Aau74840 Human Hsk

26 664.5 26.8 1031 7 ADB67093
27 660.5 26.6 975 2 AAW72746
28 660.5 26.6 975 4 ABB63485
29 660.5 26.6 975 7 ADB67088
30 658 26.5 757 4 AAU19569
31 658 26.5 757 5 ABP51294
32 656.5 26.4 829 3 AAG31117
33 656.5 26.4 934 3 AAG31116
34 656 26.4 796 7 ADE5349
35 656 26.4 1388 5 AAE14400
36 656 26.4 1388 5 AAU79590
37 656 26.4 1388 6 ABR48222
38 656 26.4 1388 7 ADB80468
39 656 26.4 1388 7 ADC35116
40 654.5 26.4 963 4 AAM78880
41 654.5 26.4 963 7 ADB67089
42 653.5 26.3 963 7 ADB67091
43 652 26.3 411 2 AAW72745
44 652 26.3 441 2 AAW72744
45 652 26.3 1362 5 AAU74840

CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2633 AA;

Query Match 100.0%; Score 2483; DB 4; Length 2633;
Best Local Similarity 100.0%; Pred. No. 7e-189;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEEGAVACVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDRVFGHNETTKN 60
DB 2 AEEGAVACVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDRVFGHNETTKN 61
QY 61 VYEEIAAPIIDSALIQYNGTIFAYGQTASGKTYTMMGSEDLHGVIPRAIHDIQKIKKFP 120
DB 62 VYEEIAAPIIDSALIQYNGTIFAYGQTASGKTYTMMGSEDLHGVIPRAIHDIQKIKKFP 121
QY 121 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVYTSMAK 180
DB 122 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVYTSMAK 181
QY 181 ITKGEKSRHYGETKMNQSSRSHTIFRMILESREKGEPSNCEGSKVSHNLNVLDSGER 240
DB 182 ITKGEKSRHYGETKMNQSSRSHTIFRMILESREKGEPSNCEGSKVSHNLNVLDSGER 241
QY 241 AAOQTGAAGVRLKEGNCINRSLFILGOVVKLSDQGVGFNYRDSKLTILQNSLGNPK 300
DB 242 AAOQTGAAGVRLKEGNCINRSLFILGOVVKLSDQGVGFNYRDSKLTILQNSLGNPK 301
QY 301 TRIICTITPVSFDETLTALQASTAKYMNTPYNEVSTDEALKRYRKEIMDLKQLEE 360
DB 302 TRIICTITPVSFDETLTALQASTAKYMNTPYNEVSTDEALKRYRKEIMDLKQLEE 361
QY 361 VSLTRAQAMEKDQLAQLLEEKDLLQVONEKIENLTMVLTSSSLTIQBELKAKRKRV 420
DB 362 VSLTRAQAMEKDQLAQLLEEKDLLQVONEKIENLTMVLTSSSLTIQBELKAKRKRV 421
QY 421 TWCLGKINKMKNNSYADQFNPTNITTKTHKLSINLLREIDSVCSDESDFNTLDTLSE 480
DB 422 TWCLGKINKMKNNSYADQFNPTNITTKTHKLSINLLREIDSVCSDESDFNTLDTLSE 481
QY 481 IEMNPAT 487
DB 482 IEMNPAT 488

RESULT 2
AAM39097
ID AAM39097 standard; protein; 2663 AA.
XX
AC AAM39097;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2242.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.

XX WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSR-) HYSRQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
XX N-PSDB; AAI58253.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
XX as central nervous system injuries.
XX
XX Example 4; SEQ ID NO 2242; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
XX encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders. Note: The sequence data for this patent did not form
XX part of the printed specification
XX
XX Sequence 2663 AA;

Query Match 100.0%; Score 2483; DB 4; Length 2663;
Best Local Similarity 100.0%; Pred. No. 7.2e-189;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEEGAVACVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDRVFGHNETTKN 60
DB 2 AEEGAVACVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDRVFGHNETTKN 61
QY 61 VYEEIAAPIIDSALIQYNGTIFAYGQTASGKTYTMMGSEDLHGVIPRAIHDIQKIKKFP 120
DB 62 VYEEIAAPIIDSALIQYNGTIFAYGQTASGKTYTMMGSEDLHGVIPRAIHDIQKIKKFP 121
QY 121 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVYTSMAK 180
DB 122 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVYTSMAK 181
QY 181 ITKGEKSRHYGETKMNQSSRSHTIFRMILESREKGEPSNCEGSKVSHNLNVLDSGER 240
DB 182 ITKGEKSRHYGETKMNQSSRSHTIFRMILESREKGEPSNCEGSKVSHNLNVLDSGER 241
QY 241 AAOQTGAAGVRLKEGNCINRSLFILGOVVKLSDQGVGFNYRDSKLTILQNSLGNPK 300
DB 242 AAOQTGAAGVRLKEGNCINRSLFILGOVVKLSDQGVGFNYRDSKLTILQNSLGNPK 301

QY 301 TRIICITPVSFDETLTALQFASTAKYMKNTPVNEVSTDEALLKRYKEIMDLKKOLEE 360
 Db 302 TRIICITPVSFDETLTALQFASTAKYMKNTPVNEVSTDEALLKRYKEIMDLKKOLEE 361
 QY 361 VSLFTRAQAMEKDQALAEKDLLOKQVONEKIENITRMLVTSSSLTLOQLKAKKRRV 420
 Db 362 VSLFTRAQAMEKDQALAEKDLLOKQVONEKIENITRMLVTSSSLTLOQLKAKKRRV 421
 QY 421 TWCLGKINKMKNVADQFNIPNITTTKTKLSINLLREIDSVCSSESVFNSNTLDTSE 480
 Db 422 TWCLGKINKMKNVADQFNIPNITTTKTKLSINLLREIDSVCSSESVFNSNTLDTSE 481
 QY 481 IEWNPAT 487
 Db 482 IEWNPAT 488

RESULT 3
 AAM40883
 ID AAM40883 standard; protein; 2688 AA.
 AC AAM40883;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 5814.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 FN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US034263.
 XX
 PR 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-00552317.
 PR 20-JUN-2000; 2000US-00598042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
 PI Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR N-PSDB; AAI60039.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.
 XX
 PS Example 2; SEQ ID NO 5814; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
 CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification
 XX
 SQ Sequence 2688 AA;
 Query Match 95.6%; Score 2374.5; DB 4; Length 2688;
 Best Local Similarity 96.7%; Pred. No. 3.5e-180;
 Matches 474; Conservative 2; Mismatches 11; Indels 3; Gaps 3;
 QY 1 AEEGAVAVCVRVPLNSREESLGETAQVYWKTDNNVIYQVDSKGFNFDRVFGHNETTKN 60
 Db 23 AEEGAVAVCVRVPLNSREESLGETAQVYWKTDNNVIYQVDSKGFNFDRVFGHNETTKN 82
 QY 61 VYEEAAPIIDSAIOGYNGTIFAYGQTASGKTYTWMGSEDHLGVIPRA-IHDIP-OKIKK 118
 Db 83 VYEEAAPIIDSAIOGYNGTIFAYGQTASGKTYTWMGSEDHLGVIPQGQFHGFSQKIXE 142
 QY 119 -PPDREFLLRVSYMEIYNETITDLCGTQKMKPLIIRVDNVRNVVADLTEEVVYTSWA 177
 Db 143 VPDREFLLRVSYMEIYNETITDLCGTQKMKPLIIRVDNVRNVVADLTEEVVYTSWA 202
 QY 178 LKWITGKSRHYGETKMNQSRSSHTIFRMILESRKGEPSNCRGSKVSHNLVLDLAG 237
 Db 203 LKWITGKSRHYGETKMNQSRSSHTIFRMILESRKGEPSNCRGSKVSHNLVLDLAG 262
 QY 238 SERAAQTGAAGVRLKEGNCINRSLEILQGVIKKLDGQGVGGFINVRDSKLTFRILONSIGG 297
 Db 263 SERAAQTGAAGVRLKEGNCINRSLEILQGVIKKLDGQGVGGFINVRDSKLTFRILONSIGG 322
 QY 298 NPKTRIICTITPVSFDETLTALQFASTAKYMKNTPVNEVSTDEALLKRYKEIMDLKKQ 357
 Db 323 NPKTRIICTITPVSFDETLTALQFASTAKYMKNTPVNEVSTDEALLKRYKEIMDLKKQ 382
 QY 358 LEEVSLETRAQAMEKDQALAEKDLLOKQVONEKIENITRMLVTSSSLTLOQLKAKRK 417
 Db 383 LEEVSLETRAQAMEKDQALAEKDLLOKQVONEKIENITRMLVTSSSLTLOQLKAKRK 442
 QY 418 RRVTCWLGKINKMKNVADQFNIPNITTTKTKLSINLLREIDSVCSSESVFNSNTLDT 477
 Db 443 RRVTCWLGKINKMKNVADQFNIPNITTTKTKLSINLLREIDSVCSSESVFNSNTLDT 502
 QY 478 LSEIENWPAT 487
 Db 503 LSEIENWPAT 512

RESULT 4
 AAY01632
 ID AAY01632 standard; protein; 2954 AA.
 XX
 AC AAY01632;
 XX
 DT 22-JUN-1999 (first entry)
 XX
 DE Amino acid sequence of centromere-associated protein-E (CENP-E).
 KW CENP-E; centromere-associated protein-E; ATPase activity;
 KW plus end-directed microtubule motor activity; chromosome congression;
 KW microtubule binding activity; chromosome movement; mitosis;
 KW cell proliferation; tumor; metastasis; vascular malfunction;
 KW inflammatory disease; immune disease; angiogenesis; hypertension;
 KW restenosis; fungal infection; selective herbicide; fungicide;
 KW insecticide; plant growth regulator; activator; cancer cell marker.
 XX
 OS Xenopus sp.
 XX
 FN W09913061-A1.

XX 18-MAR-1999.
XX PD
XX PF 10-SEP-1998; 98WO-US019231.
XX PR 11-SEP-1997; 97US-0058645P.
XX PA (REGC) UNIV CALIFORNIA.
XX PI Wood KW, Sakowicz R, Goldstein LSB, Cleveland DW;
XX WPI; 1999-229233/19.
XX DR N-PSDB; AAX26819.
XX PT Centromere-associated protein-E and related nucleic acid.
XX PS Claim 5; Page 66-67; 77pp; English.
XX
CC The present sequence represents CENP-E (centromere-associated protein-E) of Xenopus. The protein has at least one of plus end-directed microtubule motor activity, Arpase (adenosine triphosphatase) activity and microtubule binding activity. CENP-E is the motor that powers chromosome movement toward microtubule plus ends and is essential for congression of chromosomes during mitosis. Modulators of CENP-E can thus control cell proliferation. Agents that modulate CENP-E activity are lead therapeutic, bioagricultural and diagnostic agents, e.g. for treatment of unwanted cell proliferation (typical of many examples are tumors and metastases; cell proliferation; inflammatory and immune diseases; angiogenesis; vascular malfunction; and fungal infections), also as plant-hypertension; restenosis; and fungal infections), also as plant-protection agents (selective herbicides, fungicides and insecticides) and plant growth regulators or activators for improving yields. CENP-E is also a diagnostic marker for dividing cells, including cancer cells

Query Match 63.8%; Score 1584; DB 2; Length 2954;
Best Local Similarity 64.0%; Pred. No. 7.2e-117;
Matches 320; Conservative 68; Mismatches 96; Indels 16; Gaps 6;
QY 1 AEGAVAVCVRVRLNREESLGTAVYVQVVDGSKSFNDRVFGHNETTKN 60
DB 2 SEGDAVAVCVRVRLNREESLGTAVYVQVVDGSKSFNDRVFGHNETTKN 59
QY 61 VYERIAPIIDSAIQGYNGTIFAYGQTASGKTYTMGSEDLGVIPRAIHDIPOKIKKFP 120
DB 60 IYQELVPIELSAIQGYNGTIFAYGQTSSGKTYTMGTPNSLGIIPQAIQEVFKIIQIP 119
QY 121 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRVYVADLTVEEVYVTSMALKW 180
DB 120 NREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRVYVADLTVEEVYVTSMALKW 179
QY 181 ITKGEKSRHYGETKQNRSSRSHTIFRMILESRKGEPS---NCEGSVKVSHLNLVDLAG 237
DB 180 IKKGEKSRHYGETKQNRSSRSHTIFRMILESRKGEPS---NCEGSVKVSHLNLVDLAG 239
QY 238 SERAAQTGAAGVRLKEGCNINRSLFILGQVLIKLSGQGVGFINRDSKLTFRILQNSLGG 297
DB 240 SERASQTGAAGVRLKEGCNINRSLFILGQVLIKLSGQGVGFINRDSKLTFRILQNSLGG 299
QY 298 NPKTRITITTPVSFDETLTALQFASHTAKYMNTPYNEVSTDEALKRYKEIMDLAKQ 357
DB 300 NAKTVIITITTPVSFDETLTALQFASHTAKYMNTPYNEVSTDEALKRYKEIMDLAKQ 359
QY 358 LE--EVSLETRAQAMEKDLQALLPEKDLQVQNEKLENLRLMLVTSLSLTLOQELKAK 415
DB 360 LENLESSETQAQMAKEHTQLLAEIKQLHKEKREDRIWHLTNIVVASQES-QQDQVK 418
QY 416 RRRRTWCLGKINKMKNYADQFN-----IPNTIITKTHKLSINLLRDEIVCSSESDF 471
DB 419 RRRRTWAPGKIQLNSHASGVSDFMLSRPLGNFSKAKFSDMPFPEIDDSVCTEFSDF 478
QY 472 SNTLDTLS-----EIEWNPAT 487

DB 479 DDALSMDSDNGIDAENWNLAS 498
RESULT 5
ABU53125
ID ABU53125 standard; protein; 366 AA.
XX AC ABU53125;
XX AC
XX 15-APR-2003 (first entry)
XX DT Intracellular trafficking-associated DKFZphtes3_26g22 homologue #2.
XX DE Human; gene therapy; vaccine; disease treatment; detection.
XX KW Homo sapiens.
XX OS
XX PN WC200112659-A2.
XX PD 22-FEB-2001.
XX PF 18-AUG-2000; 2000WO-IB001496.
XX PR 18-AUG-1999; 99US-0149499P.
XX PR 28-SEP-1999; 99US-0156503P.
XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX PI Wiemann S;
XX DR WPI; 2001-327840/34.
XX PT Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies.
XX PS Example III; Page 745; 1095pp; English.
XX CC This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence represents a homologue of a polypeptide described in the disclosure of the invention

Query Match 33.0%; Score 820; DB 4; Length 366;
Best Local Similarity 50.5%; Pred. No. 5.1e-57;
Matches 187; Conservative 42; Mismatches 115; Indels 26; Gaps 7;
QY 11 RVRPLNREESLGTAVYVQVVDGSKSFNDRVFGHNETTKNYYEE 64
DB 1 RCRPLNREINDGSCVQVQPPWTGYKTVNHGEGSDPHKSTTFDHFVWNCQEDVYDT 60
QY 65 IAAPIIDSAIQGYNGTIFAYGQTASGKTYTMWG-----SEDLGVIPRAIHDIPOKIKKFP 120
DB 61 VAHPIVDCCPHGYNCTIFAYGQTGSKTYTMGPGGEPDHMGIPRCCHDIFDRIDRFQ 120
QY 121 DRE---FLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRVYVADLTVEEVYVTSSEM 176
DB 121 EKDHDFVHKVSYMEIYNEEYDILLCPNQPQMKPLNIEHPNMGYPVUGCTFEHVCSDYED 180
QY 177 ALKWTITKGEKSRHYGETKQNRSSRSHTIFRMILESRKGEPSNCEGSVKVSHLNLVDIA 236
DB 181 ACHWIQGNKRNHVAATNMNDHSSRSHTIFTHVEQRHK-----QCDEHVCCHKMNLVDIA 236
QY 237 GSERAAQTGAAGVRLKEGCNINRSLFILGQVLIKLSGQV-----GGFINRDSKLTFR 290
DB 237 GSERVNRITGAEGRKLEGCNINQSLTTLGNVINALADGQTKYMYGGHGIPIYRDSKLTWL 296

Fri Aug 6 10:49:12 2004

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 677 AA;

Query Match 28.9%; Score 716.5; DB 4; Length 677;
Best Local Similarity 40.8%; Pred. No. 2.3e-48; Indels 35; Gaps 13;
Matches 189; Conservative 73; Mismatches 166;

QY 6 VAVCVVRPLNSREBSLGETAQQVYKTDNNVYQV-----DGSKSFNDFRVHGNETT 58
DB 21 VRVVVTRMDKNEISAGALSISVDKINRAITVMPKNATANEPPKTYIFDNVFDGGSNQ 80

QY 59 KNYVEIAAIIIDSAIQYNGTIFAYGQTASGKTYTMMGSEDH---LGVIPRAIHDFQK 115
DB 81 MDLYVDTARPIVDKVLGYNGTILAYGQTGTGKTYTMSGNPDSPQTKGIIPNAFHIFGH 140

QY 116 IKFPDP-RBFLRVSVWEIYNETITDLCGTQKMKPLIREDVYNNVYVADLTVEEVVYTS 174
DB 141 IAKAKENQKPLVRVSVWEIYNEEVRDL--GKOVGKSLEVKERPDIGVFKDLGSGYVHNA 199

QY 175 EMALKWITGKESRHYGETKQNRSSRSHTIFRMILESREKGFSPNCEGSKVYSHLMLVD 234
DB 200 DDLNIMRLGNKRAVGAATKQNSRSHAFITVERSELGEGD--VQHVRMGKLQVLD 257

QY 235 LAGSRAAQTGAAGVRLKEGCNINRSLFGLQVKKLSGQGVGFNYRDSKLTRELONS 294
DB 258 LAGSERQKTAQSGQRLKEATKINLSLVGLNVSALVDGK--STHPIYRNSKLTRELQDS 316

QY 295 LGGNPKTRIICTITPV--SPDETILALQFASHTAKYMKNTPYVNEVSTDEALLKRYRKEIM 352
DB 317 LGGNSKTMVCAITSPADSNYMETISTLYASRAKQIQRHMHINEEPKD-ALLRHFQBEIA 375

QY 353 DLKKQLLEE-VSLETRAQAMEKDLQALKEELKDLQKQVNEKIENLRMLVTSLSLTQQ- 410
DB 376 RLKQLEEGDSLEEPSSE-----EBEDTADDELEAPLE-----IELESSTIQAV 421

QY 411 ELKAKRKRRTWC-LGKINKMKNVADQNFNIPNTIKTKHL 452
DB 422 EKKPKKREKTDABKEELAKRKHQEHAKTEQETLRNKL 464

RESULT 9
ADD49938
ID ADD49938 standard; protein; 1232 AA.
AC ADD49938;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human lung specific tumour antigen L 1447p.
XX
KW Human; lung cancer antigen; cytostatic; lung cancer; gene therapy;
KW vaccine; T-cell; tumour.
XX
OS Homo sapiens.
XX
PN US2003194764-A1.
XX
PD 16-OCT-2003.
XX
PF 04-APR-2002; 2002US-00116712.
XX
PR 05-APR-2001; 2001US-0282289P.
PR 05-OCT-2001; 2001US-0327511P.
XX (CORI-) CORIXA CORP.
PA Bangur CS, Switzer A;
XX
PI
XX

DB 61 DEMAKHIVHACQGFNGTIFAYGQTSKGKTYTMMGDEQNPQVWVLAKEIFQOISETER 120
QY 123 EPLLRSVWEIYNETITDLCGTQKMKPLIREDVYNNVYVADLTVEEVVYTSSEM-ALKWI 181
DB 121 DFLLRVGYEIEYNEKIYDLL--NKNQDLKIHESGNGIYVNVN--CEECHITSEVDLLRL 176

QY 182 TKGESRHYGETKQNRSSRSHTIFRMILESREKGFSPNCEGSKVYSHLMLVDLAGSERA 241
DB 177 CLGNKERTVGETNWNERSRSHAFILIIISR-KSDHSD-DDAVIQSVLNLVDLAGSERA 234

QY 242 AQTGAAGVRLKEGCNINRSLFGLQVKKLSGQGVGFNYRDSKLTRELONS LGGNPKT 301
DB 235 DQTGARGARLKEGCHINKSLFLSNVKSLSNADNRFTNYRDSKLTRELQASLGNAFT 294

QY 302 RIICITPVSPDETILALQFASHTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKQLEEV 361
DB 295 SIITCTKPSIMEESQTSLSATRAKKIRIKPQVNEWSDATMMKRLEREIKVLKDLAE- 353

QY 362 SLETRAQAMEKDLQALKEELKDLQKQVNEKIENLRM-----LVTSSSLTLOQELKA 414
DB 354 -----EER---KQENQKVEHLERQIKHDMHKIICGHSLS---DKG 388

QY 415 KRKRRTWC-----LGKINKMKNVADQF---NIPNTIKTKHL-----SI 454
DB 389 QQRKRRTWCPTAGSHLELAETGTEDRIDQPKVSHLPKPVFFHTSNAGKRWDNIPKTI 448

QY 455 NILREIDESVCSRDVPS 472
DB 449 NILGSLD--IGTESNSIS 464

RESULT 8
ABB65183
ID ABB65183 standard; protein; 677 AA.
XX
AC ABB65183;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 22341.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR N-PSDB; ABL09286.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
PS Disclosure; SEQ ID NO 22341; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention

DR WPI: 2003-844452/78.
 DR N-PSDB; ADD49936.
 XX New isolated polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cancer, particularly lung cancer.
 PS Claim 1; SEQ ID NO 670; 250pp; English.
 CC The invention relates to an isolated polynucleotide (a) comprising any of
 CC the 666 fully defined nucleotide sequences appearing as ADD49269 -
 CC ADD49936, ADD49936 and ADD49938, complements of (a); sequences of at
 CC least 20 contiguous residues of (a); sequences that hybridise to (a)
 CC under highly stringent conditions; sequences having at least 75 or 90%
 CC identity to (a); or degenerate variants of (a). Also included are an
 CC isolated polypeptide (b) (comprising: sequences encoded by the new
 CC polynucleotide; any of the 4 amino acid sequences fully defined in the
 CC specification; or sequences having at least 70 or 90% identity to the
 CC sequence in (a) or (b)), an expression vector comprising the above
 CC polynucleotide operably linked to an expression control sequence, a host
 CC cell transformed or transfected with the above expression vector, a host
 CC isolated antibody, or its antigen-binding fragment, that specifically
 CC binds to the above polypeptide, an oligonucleotide that hybridises to the
 CC above-mentioned nucleotide sequences under highly stringent conditions; a
 CC fusion protein comprising at least one polypeptide cited above, detecting
 CC the presence of a cancer in a patient (comprising: obtaining a biological
 CC sample from the patient; contacting the biological sample with a binding
 CC agent that binds to the polypeptide, or with the oligonucleotide cited
 CC above; detecting in the sample an amount of the polypeptide that binds to
 CC the binding agent, or an amount of a polynucleotide that hybridises to
 CC the oligonucleotide; and comparing the amount of polypeptide, or
 CC polynucleotide that hybridises to the oligonucleotide, to a predetermined
 CC cut-off value and then determining the presence of a cancer in the
 CC patient), a method for stimulating and/or expanding T-cells specific for
 CC a tumour protein (comprising contacting T-cells with the above
 CC polypeptide, polynucleotide or antigen-presenting cells that express the
 CC polynucleotide, under conditions and for a time sufficient to permit the
 CC stimulation and/or expansion of T-cells), an isolated T-cell population
 CC comprising T-cells prepared by the method, a composition comprising a
 CC first component selected from physiological carriers and
 CC immunostimulants, and a second component selected from the above
 CC polypeptide, polynucleotide, antibody, fusion protein, T-cell population
 CC and antigen-presenting cells that express the above polypeptide,
 CC stimulating an immune response in a patient (comprising administering to
 CC the patient the above composition) treating lung cancer in a patient
 CC (comprising administering to the patient the above composition and a
 CC diagnostic kit (comprising: at least one oligonucleotide cited above; or
 CC at least one antibody cited above and a detection reagent, where the
 CC detection reagent comprises a reporter group). The composition and
 CC methods are useful in diagnosing, preventing and treating cancer,
 CC particularly lung cancer. The present sequence is a lung cancer-
 CC associated antigen of the invention.
 XX Sequence 1232 AA;
 SQ
 Query Match 28.7%; Score 713.5; DB 7; Length 1232;
 Best Local Similarity 35.1%; Pred. No. 9.5e-48;
 Matches 182; Conservative 86; Mismatches 191; Indels 59; Gaps 11;
 QY 6 VAYCVVRPLNRSSEISG-ETAQVYWKTDNNVYQVDSKSENFDFVHGNETHKNVYEE 64
 DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 10 VRVALRCPLVPKSEISGCMCLSFVGPGEQVVGVD--KSFYDFVDFDSTQEVEFNT 67
 QY 65 IAPITDSIAQYNGTIFAYGQTASGKTYTMMG-----SEHLGVIPRAIHDFOKIK 117
 DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 68 AVAPLKGVPKYNATVAYGQTGSGKTYSMGAYTAQEENETVGVIPVLIQLLKEID 127
 QY 118 KFPDFRFLRVSMYELNETITLLCGTQKMKELIREDNRNRYVADLTVEVYVYSEMA 177
 DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 128 KGSDFFTKLVSYELTYNEIIDLCPSEKAKQINIREDPKEGKIVGLTEKTVLVALDT 187
 QY 178 LKWTIGKESRHYGKMNORSRSHITPFMILESREKGEPSNCEGSKVSHLVLDLAG 237
 DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 188 VSLCEGNNRSRTVASTAMNSQSSRSHAITISLEQRKSD----KNSFSRSLHLVDLAG 243

QY 238 SERAAQTGAAGVRLKEGCMNIRSLFILQVVIKLSGQVGGFINYRDSKILTRILQNSLGG 297
 DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 244 SERQKTKAEGRLKEGINIRGLLCLGNVISALGDGKGGFVPYRDSKILTRILQNSLGG 303
 QY 298 NPKTRIICITTPV--SFDLTLALQFASTAKYMKNTFYNEVSTDEALLKRYKEIMDLK 355
 DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 304 NSHTLMIACVSPADSNSLEETLRYADARKIKKPIVN-----IDPQTAELNHLK 355
 QY 356 KOLEEVS-----ETRAQAMKDXLAQLLEEKDLQKQVNEKIENL 396
 DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 356 QOVQQLQVLLQAHGGTLPSSITVPESENLSIMEKNQ--SLVEENEKLSRGSEAGQT 413
 QY 397 TWML---VTSSSLITLQOELKAKRKRVTWCLGKINK-----MKNSYADQFNPTNTTKT 449
 DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 414 AQMLRIILTEQANFKMAKLELRQHAACKLDLQKLVELTDELQELKENVEIICNLQQLI 473
 QY 450 HKLSINLLREIDESVCSDSDFNTLDTLSELEWNPAT 487
 DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 474 TQLS-----DETACMAAIDTAVEQAEQVETSPET 504
 RESULT 10
 ADD18924
 ID ADD18924 standard; protein; 1232 AA.
 AC ADD18924;
 XX
 DT 15-JAN-2004 (first entry)
 DE Human disease related protein SeqID413.
 KW human; disease state; cytostatic; antiinflammatory; ophthalmological;
 KW antiarteriosclerotic; vulnary; gene therapy;
 KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;
 KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;
 KW glucose transport; catecholamine synthesis;
 KW nitric oxide synthesis; cancer; ischaemic condition; iron transport;
 KW reinnopathy; neonatal stress; pre-eclampsia; atherosclerosis;
 KW inflammatory condition; wound healing.
 XX Homo sapiens.
 OS
 XX WO2003018621-A2.
 FN
 XX 06-MAR-2003.
 PD
 XX 23-AUG-2002; 2002WO-GB003892.
 PF
 XX 23-AUG-2001; 2001GB-00020558.
 PR
 XX 05-OCT-2001; 2001GB-00024037.
 XX
 PA (OXFO-) OXFORD BIOMEDICA UK LTD.
 XX
 XX Kingsman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;
 PI
 XX WPI; 2003-230046/28.
 DR
 XX N-PSDB; ADD18925.
 XX
 PT New substantially purified polypeptide, useful for diagnosing or treating
 PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion
 PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or
 XX wound healing.
 PS Claim 25; SEQ ID NO 413; 424pp; English.
 XX
 CC This invention relates to novel human genes and gene product which are
 CC implicated in certain disease states. Compounds which modulate the
 CC proteins of the invention may have cytostatic, antiinflammatory,
 CC ophthalmological, antiarteriosclerotic or vulnary activities. The
 CC sequences of the invention may be useful for gene therapy. The invention
 CC may be useful for diagnosing or treating a hypoxia-regulated condition,
 CC such as tumorigenesis, angiogenesis, apoptosis, inflammation,

RESULT 12
ADCL0190
ID ADC10190 standard; protein; 1029 AA.
XX AC ADC10190;
XX DT 18-DEC-2003 (first entry)
XX DE Human NOVX polypeptide SEQ ID NO: 212.
XX KW Cytostatic; antidiabetic; anorectic; cerebroprotective; neuroprotective;
KW antiinflammatory; gene therapy; antisense therapy; thrombolytic; NOVX;
KW pathology; cancer; diabetes; obesity; endocrine disorder; CNS disorder;
KW inflammatory disorder; chromosome mapping; tissue typing;
KW predictive medicine.
XX OS Homo sapiens.
XX PN WO2003000842-A2.
XX PD 03-JAN-2003.
XX PF 04-JUN-2002; 2002WO-US017443.
XX PR 04-JUN-2001; 2001US-0295607P.
XX PR 04-JUN-2001; 2001US-0295661P.
XX PR 06-JUN-2001; 2001US-0296404P.
XX PR 06-JUN-2001; 2001US-0296418P.
XX PR 07-JUN-2001; 2001US-0296575P.
XX PR 11-JUN-2001; 2001US-0297414P.
XX PR 12-JUN-2001; 2001US-0295573P.
XX PR 12-JUN-2001; 2001US-0297567P.
XX PR 14-JUN-2001; 2001US-0298285P.
XX PR 15-JUN-2001; 2001US-0298528P.
XX PR 18-JUN-2001; 2001US-0299133P.
XX PR 19-JUN-2001; 2001US-0299230P.
XX PR 21-JUN-2001; 2001US-0299949P.
XX PR 22-JUN-2001; 2001US-0300177P.
XX PR 26-JUN-2001; 2001US-0300883P.
XX PR 28-JUN-2001; 2001US-0301530P.
XX PR 28-JUN-2001; 2001US-0301550P.
XX PR 03-JUL-2001; 2001US-0302951P.
XX PR 31-JUL-2001; 2001US-0308909P.
XX PR 14-SEP-2001; 2001US-0322297P.
XX PR 25-SEP-2001; 2001US-0324669P.
XX PR 03-DEC-2001; 2001US-0337477P.
XX PR 14-DEC-2001; 2001US-0341562P.
XX PR 21-FEB-2002; 2002US-0358656P.
XX PR 21-FEB-2002; 2002US-0359122P.
XX PR 22-FEB-2002; 2002US-0358978P.
XX PR 22-FEB-2002; 2002US-0359034P.
XX PR 22-FEB-2002; 2002US-0359035P.
XX PR 27-FEB-2002; 2002US-0359121P.
XX PR 27-FEB-2002; 2002US-0359964P.
XX PR 01-MAR-2002; 2002US-0360859P.
XX PR 12-MAR-2002; 2002US-0363430P.
XX PR 12-MAR-2002; 2002US-0363678P.
XX PR 10-APR-2002; 2002US-0371346P.
XX PR 10-MAY-2002; 2002US-0379444P.
XX PR 04-JUN-2002; 2002US-00379444.
XX PA (CURA-) CURAGEN CORP.
XX PI Agee ML, Anderson DW, Berghs C, Casman SU, Catterton E;
PI Dipippo VA, Edinger SR, Eisen A, Ellerman K, Gangolli EA;
PI Gerlach VL, Gorman I, Guo X, Herrmann JL, Hjal T, Ji W, Kekuda R;
PI Khramtsov NV, Li L, Malyankar UM, Miller CE, Millet I;
PI Ort T, Padigar M, Patturajan M, Pena CEA, Rastelli L, Rieger DK;
PI Rothenberg ME, Shency SG, Shimkets RA, Smithson G, Spaderna SK;
PI Spytek KA, Stone DJ, Vernet CAM, Zhong H, Zhong M, Alsobrook JP;
PI Burgess CE, Lepley DM;

XX WPI: 2003-210149/20.
DR N-PSDB; ADC10189.
XX
PT New isolated NOVX polypeptides and nucleic acid molecules useful for
PT treating, preventing and diagnosing pathological conditions with NOVX-
PT associated disorders, such as cancer, obesity, diabetes and inflammatory
PT or CNS diseases.
XX
PS Claim 1; SEQ ID NO 212; 772pp; English.
XX
CC The invention relates to novel isolated polypeptides, mature form of the
CC polypeptide, a sequence that is 95% identical to the polypeptide or the
CC polypeptide comprising one or more conservative substitutions. The NOVX
CC polypeptide is useful for treating or preventing a pathology associated
CC with the polypeptide e.g. disorders associated with aberrant expression
CC or activity of the polypeptide, such as cancer, diabetes, obesity, and
CC endocrine, CNS and inflammatory disorders. They can also be used in
CC various detection and screening assays, chromosome mapping, tissue typing
CC and predictive medicine. This sequence corresponds to one of the
CC polypeptides of the invention.
XX
SQ Sequence 1029 AA;
Query Match 28.6%; Score 710.5; DB 7; Length 1029;
Best Local Similarity 36.5%; Pred. No. 1.3e-47;
Matches 190; Conservative 80; Mismatches 176; Indels 75; Gaps 16;
QY 5 AVAVCVVRPLNSREESLGETAQVYKTD-----NNVIYQVDSKSFNDRVPHGNE 56
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 5
5 AVKVVRRCRPMQORREL--RCQPVTVDCARAQCQIQNGAADPEPKQTFDGAHVHDH 62
QY 57 TTKNVEHIAPIIDSAIQYNGTIFAYGQTASGKTYTMGSED---HLGVIPRAIHDF 113
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 63
63 VTEQIYNEIAYPLVEGVTGYNGTIFAYGQTGSGKFTMQGLPDPSPQRGIIPRAPEHVF 122
QY 114 QKIKFPDREFLLRVSYMEIYNETITDLCGQKMKPLIREDVNRNVVADLTBRVVYT 173
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 123
123 ESVQCAENTKFLVRASYLEIYNEVDRLDGLGATKQK-LELKEHPEKGVTVKGLSMITVHS 181
QY 174 SEMALKWITGKSRHYGETKMNORSRSHITFRMILESR---EKGEPSNCSGVKVSHL 230
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 182
182 VAQCEHIMETGWNRSVGVTLNKKDSRSHSIFTISIEMSAVDERG----KDLRAGKL 236
QY 231 NLVDLAGSERRAQTCAAGVRLKEGNCINRSLFILGVIKKLSDGQGVGFINTRDSKLTRI 290
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 237
237 NLVDLAGSERSQSGTATGERLKEATKINLSLALGNVISALVDGRC-KHVPYRDSKLTEL 295
QY 291 LQNSLGNPKTRIICITIPV--SPDETTLAQFASTAKYMKNTPVVNEVSTDEALLKRYR 348
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 296
296 LQDSLGGNTKTLMVACLSPADNNYDETTLSTLYANRKNRKNRKNRKNRKNRKNRKNRKN 354
QY 349 KEIMDLKKOLEVSLFTRAQAMEKDQLAQLL-----EEKDLLQKV----- 388
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 355
355 BEIKKILKAIL-----TQMSPSLSALLSRQVPPVQVEEKLLPQPVLIQHDMEAK 406
QY 389 ---QNEKIENITRLMTVSSSLTLOELKAKKRRVTWCLGKINKNKNKNYADQFNIPNI 445
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 407
407 QLIRREYEEERLARL---KADYKAEQESFARLEEDT-----AMRNSYDRLSTLEENL 456
QY 446 TTKTKHLSINLREIDESVCSSESVFNTLDTLSIEWNPA 486
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 457
457 RKETEAV-----LQGVLYKAEVMSRA-EFASAEYPPA 489
RESULT 13
ABG70993
ID ABG70993 standard; protein; 1234 AA.
XX AC ABG70993;
XX DT 10-DEC-2002 (first entry)
XX

DE Human Hekif4b protein.

XX Microtubule stimulated ATPase; cellular proliferation; cancer; AIDS;

KW hyperplasia; restenosis; cardiac hypertrophy; immune disorder; HIV;

KW inflammation; autoimmune disease; arthritis; graft rejection; psoriasis;

KW inflammatory bowel disease; proliferation; medical procedure; surgery;

KW human immunodeficiency virus; acquired immunodeficiency syndrome;

KW angioplasty; human; Hekif4b; kinesin family.

XX

OS Homo sapiens.

XX

XX

PH Key Location/Qualifiers

FT Misc-difference 850

FT /note= "Encoded by CGBCA"

XX

XX USG440684-B1.

XX

XX 27-AUG-2002.

XX

XX 12-JUN-2000; 2000US-00592054.

XX

XX 12-JUN-2000; 2000US-00592054.

XX

XX (CYTO-) CYTOKINETICS INC.

XX

XX Beraud C, Finer JT, Sakowicz R, Wood KW;

XX

XX WPI; 2002-711529/77.

XX

XX N-PSDB; ABS55163.

XX

XX Screening for modulators of target protein having microtubule stimulated

PT ATPase activity e.g. kinesin family of protein, useful for treating

PT cancer, psoriasis, arthritis, human immunodeficiency virus (HIV)

PT infection.

XX

XX Disclosure; Fig 8; 34pp; English.

XX

XX The present invention relates to a new method of screening modulators of

CC target protein with microtubule stimulated ATPase activity. The method

CC involves contacting the target protein with an agent at 1st and 2nd

CC concentrations and determining the level of activity (e.g. binding or

CC ATPase activity) of target protein, where a difference between levels of

CC activity of target protein contacted with 1st and 2nd concentrations of

CC an agent indicates that an agent modulates activity of target protein.

CC The invention can be used for screening for modulators of target protein

CC having microtubule stimulated ATPase activity. The compounds identified

CC by method of the invention are useful for treating cellular proliferation

CC including cancer, hyperplasias, restenosis, cardiac hypertrophy, immune

CC disorders and inflammation. The compounds identified by the method are

CC also useful for treating autoimmune disease, arthritis, graft rejection,

CC inflammatory bowel disease, proliferation induced by medical procedures,

CC e.g. surgery, angioplasty etc. The compounds are also useful for treating

CC psoriasis. The compounds are useful for inhibiting human immunodeficiency

CC virus (HIV) and thus treating acquired immunodeficiency syndrome (AIDS).

CC The present amino acid sequence represents the human Hskif4b (kinesin

CC family) protein as described in the invention

XX

XX Sequence 1234 AA;

XX

Query Match 28.4%; Score 705.5; DB 5; Length 1234;

Best Local Similarity 34.9%; Pred. No. 4,2e-47;

Matches 182; Conservative 93; Mismatches 181; Indels 65; Gaps 14;

QY 6 VAVCVVRPLNREESLG-ETAQVYMKTDNNVIYQVDSKSFNDFRVFHNQETTKNVEE 64

DB 10 VRVALRCRPLVPKEISEGQMCGLSFVPGTQVVVGTG--KSFTYDFVDPCTEQEVENK 67

QY 65 TAAPIDSAIQYNGTIFAYGTAGSKTYTMG-----SEDLGVIPRAHIDIFOKIK 117

DB 68 AVAPLIKIGFKYNATVLAAYGTGSGKTYSMGGAYTAQENEPYTGIIIPVQLLFKEID 127

QY 118 KFPDFEFLRVSMELNYETITDLGCTGCKMPLIREDVNRNVVADLTBEWVYTSEMA 177

Db 128 QKSDFEFTLVSYLEIYNBEEILDLCPSREKAQINIREDPKEGKIIVGLTEKTVLVALDT 187

QY 178 LKWITGKSRHYGETHYKQNRSSRSHITIFRMILESEKCEPSNCEGSKVKVSHLNLYDLAG 237

Db 188 VSCLEQGNNSRTVASTAMNSQSSRSHAFTI---SLEQKKSKDNSSFR-SKLHLVDLAG 243

QY 238 SERAAQTGAAGVRLKGCNINRSIFILGOVKKLSQGVGFNYRDSKLTIRLQNSLGG 297

Db 244 SERQKTKAEGDRLEKGININRGLLGNVISALGDDKGSFVPYRDSKLTIRLQDSLGG 303

QY 298 NPKTRIITITPV--SFDETLTALQFASHTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLK 355

Db 304 NSHTLMTACVSPADSNLEETLSTLRVADRARKINKKEIVN-----IDPHTAELNHLK 355

QY 356 KQLEEVSL-----ETRAQMEKQDLAQLLEEKD-----LLQKVQNEK 392

Db 356 QVQVQQLQVILLQAHGGLPGSINAEPSNLQSLMEKNQ--SLVEENKLSRCLSKAAGQT 413

QY 393 IENIRMLVTSSSLTLOELKAKRK--RRVTWCIGLKINK---MKNSNYADQFNIPNTIT 446

Db 414 AQMLERILITEQ---VNEKLNKLEELRQHAACKLDLQKLVETLEDQELKENVEICNLQ 470

QY 447 TKTHKLSINLLREIDESVCSFVSNLTDLTSLSEIWNPAT 487

Db 471 QLITQLS-----DETVACTAAADTAVBEEAQVETSPET 504

RESULT 14

AAG31112

ID AAG31112 standard; protein; 1034 AA.

XX

XX AAG31112;

XX

DT 17-OCT-2000 (first entry)

XX

XX Arabidopsis thaliana protein fragment SEQ ID NO: 37309.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-00301439.

XX

PR 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 28-APR-1999; 99US-0130891P.

PR 30-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 04-MAY-1999; 99US-0132407P.

PR 05-MAY-1999; 99US-0132484P.

PR 06-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132486P.

PR 07-MAY-1999; 99US-0132487P.

PR 11-MAY-1999; 99US-0132863P.

PR 14-MAY-1999; 99US-0134256P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 18-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0146389P.
PR 04-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 08-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159291P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.

```
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      28.4%; Score 705; DB 3; Length 1034;
Best Local Similarity 34.3%; Pred. No. 3.6e-47;
Matches 196; Conservative 96; Mismatches 148; Indels 132; Gaps 18;

QY 1 AEEGAVCVVRRLN-----SREES-----LGETA---QVVMKTDN 34
Db 7 SERDSISVTVRFLRVARSLAMDQIYCRDYSFHVDAIGVNSLLGEVQRDEVAWPDG 66
QY 35 NVI--YQVDGSKSFNFORVHGNETHKNNVYEEIAPIIDSIAIOGNYNGTIFAYGQTASGKT 92
Db 67 DTLVRHEYNPLTAFADKVFQPGQATTIDYDVAARPVVKAAMEGVNGTVFAYGVTSSGKT 126
QY 93 YTM-----MGSEDLHLGVIPRAIHDF 113
Db 127 HTMVRVLKKNPYTPPSLVFLQNTCAMNWDLFFNLVPLWLLGDQSPGIIPLAIKDVF 186
QY 114 QKIKKF-----PREFLLRYSMEIYNETITDLCGQKMKPLIREDVNRNVVADLT 167
Db 187 SIIQDVSLNGTGPREFLLRYSLEIYNEVINDLLDPTG--QNLEVRD-SQGYVEGIX 243
QY 168 EEVVYTSMAKWTIKGSKRHYGETKMNQRSSRSHITFRMILESRKEKPEPSNCEGSVKV 227
Db 244 EEVVLSPGHALSFIAAGEHRRHVGNNFNLLSSRSHITFTLMVSSATGDEYD---GVIF 300
QY 228 SHLMVLVDLAGERRAQTGAAGVRLKEGCNINRSLFILQGVIKKLSGQGVGFVINYRDKL 287
Db 301 SQLNLIDLAGE-SKSTETTGLRKEGSIYKSLTLTGVIGKLESGK-ATHIPYRDSKL 358
QY 288 TRILONSGGNPKRIICTIPV--SEDETITAFOPASTAKYMKNTPVYNEVSTDEALLK 345
Db 359 TRLQSSLSGSHGVSLICTIITPASSSEETHNTLAFASRAKSIETIYASRNQIIDEKSLIK 418
QY 346 RYRKEIMDLKQLEEV-----SLETRAQAMEKDQ--LAQLLEKDLKQVQNEK 392
Db 419 KYQRIISTIKLELDQLRRGMLVGVSHLEMSLKQLEEGQVKMQSRLEEEEAKAALMSR 478
QY 393 IENLRMLVTSSSLTLOQLKAKRRVTVCLGKINKMKNSNYADQFNIPITTKTKHL 452
Db 479 IQKLTKLIVST-----KNSTPGYSGDIP-----THQR 506
QY 453 SINLREID-EGVCSSESDFV---SNTLDTLSE 480
Db 507 SLGAKDDKFDLSLLESNIGLSPSSTLALLSE 538

RESULT 15
ID AAG31111 standard; protein; 1069 AA.
XX AC AAG31111;
XX DT
XX DT
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 37308.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN
PN EF1033405-A2.
```


Fri Aug 6 10:49:12 2004

Db	454	KYOREISTKLELDQRRGMLVGVSHHELSLKQQLBECQVMQSRLEBEEEAALMSR	513
QY	393	IENLTRMLVTSSLTQQELKAKRRVRTWCLGKINKMKNSNYADQFNIPNTTTHKL	452
Db	514	IQKTKLILVST-----KNSIPGYSGDIP-----THQR	541
QY	453	SINLREID-BSVCSESDVF---SNTLDTLSE	480
Db	542	SLSAGKDDKFDLLESNDLGSPTSLLISE	573

Search completed: July 29, 2004, 09:35:15
Job time : 30.3136 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:30:02 ; Search time 8.62775 Seconds
(without alignments)
2914.068 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_488

Perfect score: 2483

Sequence: 1 AEEGAVAVCVVRPLNSRRE.....SDVFSNTLDTLSIEWNPAT 487

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2483	100.0	2662	4	US-09-595-684B-31
2	1584	63.8	2954	4	US-09-150-867-1
3	705.5	28.4	1234	4	US-09-592-054-8
4	704	28.4	1231	4	US-09-595-684B-23
5	703	28.3	1066	3	US-09-541-782-8
6	703	28.3	1066	4	US-09-723-820-8
7	703	28.3	1066	4	US-10-270-085-8
8	702	28.3	522	4	US-09-592-054-4
9	696.5	28.1	1232	4	US-09-592-054-2
10	695	28.0	473	4	US-09-592-054-6
11	664.5	26.8	1031	4	US-09-914-259-24
12	660.5	26.6	975	4	US-09-914-259-19
13	656	26.4	1388	4	US-09-572-191-2
14	656	26.4	1388	4	US-09-723-262-2
15	656	26.4	1388	4	US-09-723-219-2
16	654.5	26.3	963	4	US-09-914-259-20
17	653.5	26.3	963	4	US-09-914-259-22
18	652	26.3	411	2	US-08-713-815A-4
19	652	26.3	441	2	US-08-713-815A-3
20	642.5	25.9	967	4	US-09-914-259-21
21	640	25.8	957	4	US-09-914-259-16
22	638.5	25.7	1032	4	US-09-914-259-26
23	636.5	25.6	956	4	US-09-914-259-17
24	635.5	25.6	1027	4	US-09-914-259-27
25	632.5	25.5	1056	4	US-09-595-684B-29
26	632.5	25.5	1057	4	US-09-541-782-10
27	632.5	25.5	1057	4	US-09-723-820-10

28	632.5	25.5	1057	4	US-10-270-085-10	Sequence 10, Appl
29	632	25.5	409	4	US-09-572-191-6	Sequence 6, Appl
30	632	25.5	409	4	US-09-723-262-6	Sequence 6, Appl
31	632	25.5	409	4	US-09-723-219-6	Sequence 6, Appl
32	628	25.3	815	4	US-09-914-259-18	Sequence 18, Appl
33	624.5	25.2	575	4	US-09-724-519-8	Sequence 8, Appl
34	624.5	25.2	575	4	US-09-592-037-8	Sequence 8, Appl
35	624.5	25.2	575	4	US-09-428-156B-8	Sequence 8, Appl
36	623.5	25.1	1057	4	US-09-428-156B-8	Sequence 8, Appl
37	621.5	25.0	935	4	US-09-914-259-25	Sequence 25, Appl
38	613.5	24.7	513	4	US-09-724-519-6	Sequence 6, Appl
39	613.5	24.7	513	4	US-09-592-037-6	Sequence 6, Appl
40	613.5	24.7	513	4	US-09-428-156B-6	Sequence 6, Appl
41	611.5	24.6	1111	4	US-09-914-259-28	Sequence 28, Appl
42	610	24.6	928	4	US-09-914-259-23	Sequence 23, Appl
43	608	24.5	1053	4	US-09-724-519-2	Sequence 2, Appl
44	608	24.5	1053	4	US-09-592-037-2	Sequence 2, Appl
45	603	24.3	375	4	US-09-572-191-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; TITLE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: PRT
; ORGANISM: Human
US-09-595-684B-31

Query Match 100.0%; Score 2483; DB 4; Length 2662;
Best Local Similarity 100.0%; Pred. No. 8.7e-208;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AEEGAVAVCVVRPLNSRREISLGETAQVYWKTDNNVIYVDGSKSFNDFVPHGNETTKN 60	
DB	2	AEEGAVAVCVVRPLNSRREISLGETAQVYWKTDNNVIYVDGSKSFNDFVPHGNETTKN 61	
QY	61	VVEETAAPIIDSAIQYNGTIFAYGTASGKTYTWMGSEDLGLVTPRAIHDFQKIKKFP 120	
DB	62	VVEETAAPIIDSAIQYNGTIFAYGTASGKTYTWMGSEDLGLVTPRAIHDFQKIKKFP 121	
QY	121	DREFTLLRYSYMEIYNETITDLCCGQKMKPLIREDVNRNVYVAULTTEVYVWTSEMAK 180	
DB	122	DREFTLLRYSYMEIYNETITDLCCGQKMKPLIREDVNRNVYVAULTTEVYVWTSEMAK 181	
QY	181	ITKGEKSRHYGTQKNQSSRSHTIFRMLLSREKGEPSNCEGSKVSHNLVDLAGSR 240	
DB	182	ITKGEKSRHYGTQKNQSSRSHTIFRMLLSREKGEPSNCEGSKVSHNLVDLAGSR 241	
QY	241	AAQTGAAGVRLKEGNCINRSLFILQVVKKLSGQGVGFINRYRDSKLTRELQNSLGGNPK 300	
DB	242	AAQTGAAGVRLKEGNCINRSLFILQVVKKLSGQGVGFINRYRDSKLTRELQNSLGGNPK 301	

```

QY 301 TRIICITTPVSFDETLTALQFASAKYMKNTPPYNEVSTDEALKRYRKEIMDLKKLEE 360
Db 302 TRIICITTPVSFDETLTALQFASAKYMKNTPPYNEVSTDEALKRYRKEIMDLKKLEE 361
QY 361 VSLTRAQAMEKQDLAQLEKDLQKQVNEKIENLRMLVTSSSLTQOELKAKRRRV 420
Db 362 VSLTRAQAMEKQDLAQLEKDLQKQVNEKIENLRMLVTSSSLTQOELKAKRRRV 421
QY 421 TWCLGKINKMNSYADQFNIPNTITTKTHKLSINLAREIDESVCSSEDFNTLDTLSE 480
Db 422 TWCLGKINKMNSYADQFNIPNTITTKTHKLSINLAREIDESVCSSEDFNTLDTLSE 481
QY 481 IEWNPAT 487
Db 482 IEWNPAT 488

RESULT 2
US-09-150-867-1
; Sequence 1, Application US/09150867
; Patent No. 6645748
; GENERAL INFORMATION:
; APPLICANT: Wood, Kenneth W.
; APPLICANT: Sakowicz, Roman
; APPLICANT: Goldstein, Lawrence S.B.
; APPLICANT: Cleveland, Don W.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Plus End-Directed Microtubule Motor Required for
; TITLE OF INVENTION: Chromosome Congression
; FILE REFERENCE: 18557C-000110US
; CURRENT APPLICATION NUMBER: US/09/150,867
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: US 60/058,645
; EARLIER FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2954
; TYPE: PRT
; ORGANISM: Xenopus sp.
; FEATURE:
; OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)
; OTHER INFORMATION: member of the kinesin superfamily of microtubule
; OTHER INFORMATION: motor proteins
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1)..(472)
; OTHER INFORMATION: kinesin like motor domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (473)..(2752)
; OTHER INFORMATION: rod domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (2753)..(2954)
; OTHER INFORMATION: tail domain
US-09-150-867-1

```

```

Query Match 63.8%; Score 1584; DB 4; Length 2954;
Best Local Similarity 64.0%; Pred. No. 4,7e-129; Indels 16; Gaps 6;
Matches 320; Conservative 68; Mismatches 96;

QY 1 AREGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDRVPHGNETTKN 60
Db 2 SEGDAVKVCVRPLNREQ--GDQANLQWKAGNNTISQVDTKSFNDRVFNESHSTQ 59
QY 61 VYEEAIPIDSALQSYNGTIFAYGQTASGKYTMGSEDLHGVIPRAHDFQIKKFP 120
Db 60 IYQEIAVPIRSALQSYNGTIFAYGQTSSGKYTMGTENSLGIIPQAIQEVFKIIQETP 119
QY 121 DREFLLRVSYMEIYNETITDLCCGQCKMKPLIREDVNRNVYVADLTEEVVYVTSMA 180
Db 120 NREFLLRVSYMEIYNETVTKDLCCDRKKPLIREDVNRNVYVADLTEELVMVPEHVIQW 179

```

```

QY 181 ITKGEKSRHYGETOMQSRSSHTIFRMILEBREKGEPS---NCEGSVKYSHLNLDVLAG 237
Db 180 IKKGEKNRHYGETOMKNDHSRSHTIFRMIVESDRNDPTNSENCDGAVVWSHLNLDVLAG 239
QY 238 SERAAQTGAACVRLKEGCNINRSIFILGOVKKLSDGQGVGFNYRDSKLTRILQNSLGG 297
Db 240 SERASQTGAEGVRLKEGCNINRSIFILGOVKKLSDGQAGGFNYRDSKLTRILQNSLGG 299
QY 298 NPKTRIICITTPVSFDETLTALQFASAKYMKNTPPYNEVSTDEALKRYRKEIMDLKKQ 357
Db 300 NAKTVIICITTPVSFDETLTALQFASAKYMRNTPHYNEVLDDDEALKRYRKEIMDLKKQ 359
QY 358 LE--EVSLETRAQAMEKDLAQLEKDLQKQVNEKIENLRMLVTSSSLTQOELKAK 415
Db 360 LENLESSETKAQAMAKEHTQLIAETKQLHKEDEIRIWHLTNTVWASSQES--QDQVRK 418
QY 416 RKRRTWCLGKINKMNSYADQFN---IPTNITTKTHKLSINLAREIDESVCSSEDF 471
Db 419 RKRRTWAPGKIQNSLHSGVSDFDMLSRUPGNFKKAKFSDMPSFEIDDSVCTERSDF 478
QY 472 SNTLDTLS---EIEWNPAT 487
Db 479 DDALSMDNSGIDAENLAS 498

RESULT 3
US-09-592-054-8
; Sequence 8, Application US/09592054
; Patent No. 6440684
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Finer, Jeffrey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6440684el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1016
; CURRENT APPLICATION NUMBER: US/09/592,054
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1234
; TYPE: PRT
; ORGANISM: Human
US-09-592-054-8

```

```

Query Match 28.4%; Score 705.5; DB 4; Length 1234;
Best Local Similarity 34.9%; Pred. No. 9.6e-53; Indels 65; Gaps 14;
Matches 182; Conservative 93; Mismatches 181;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDRVPHGNETTKNVEE 64
Db 10 VRVALRCRLVPKEISECQCLSFVPGETQVWGTD--KSFYDFVEDPCTEQEVFNK 67
QY 65 IAAPIIDSALQSYNGTIFAYGQTASGKYTMG-----SEDLHGVIPRAHDFQIK 117
Db 68 AVAPLIKGFYKYNATVLAYGQTSGKTYSMGAYTAQENEPYGLIIPRVQLLFKEID 127
QY 118 KFPDREFLLRVSYMEIYNETITDLCCGQCKMKPLIREDVNRNVYVADLTEEVVYVTSMA 177
Db 128 QKSDSEFTLKVSYLEIYNEEILDLCLPSREKQAQINREDPKEGIKIVGLTEKTVLVALDT 187
QY 178 LKWIITKGEKSRHYGETKNQSRSSHTIFRMILEBREKGEPSNCSGVKVSHLNLDVLAG 237
Db 188 VSCLEQGNNSRTVASTAMNSQSSRSHTAFTI---SLEQKKSKDKNSSFR-SKLHLVDLAG 243
QY 238 SERAAQTGAACVRLKEGCNINRSIFILGOVKKLSDGQGVGFNYRDSKLTRILQNSLGG 297
Db 244 SERQKTKAEGDRLKEGININRGLCLGNVISALGDDKSGSFVPYRDSKLTRILQNSLGG 303
QY 298 NPKTRIICITTPV--SFDETLTALQFASAKYMKNTPPYNEVSTDEALKRYRKEIMDLK 355

```

```
Db 304 NSHTLMIACVSPADSNLEETLSTLRVADRARKNKPIVN-----IDPHTAELNHLK 355
QY 356 KQLEEVSL-----ETRAQAMEKQDLAQLLEKD-----LQKVONEK 392
Db 356 QVQVQLQVLLQAHGGLPGSINAPESENLSLMENKQ--SLVEENEKLSRCLSKAAGQT 413
QY 393 IENLRLMTVSSSLTQOELKAKRK--RRVTWCLGKLNK-----MKSNYADQFNIPNIT 446
Db 414 AQMLERILITEQ---VNEKLNAKLELRQHAACKLDLQKLVELEDOELKENVEILCNLQ 470
QY 447 TKTHKLSINLLREIDRSVCSDESDFNTLDTLSEIENPAT 487
Db 471 QLITQLS-----DETVACTAAAIIDTAVBEAQVETSPET 504

RESULT 4
US-09-595-684B-23
; Sequence 23, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohaeshi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; and purifying human kinesins
; FILE REFERENCE: cytopo36
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 1231
; TYPE: PRT
; ORGANISM: Human
US-09-595-684B-23

Query Match 28.4%; Score 704; DB 4; Length 1231;
Best Local Similarity 35.4%; Pred. No. 1.3e-52;
Matches 183; Conservative 85; Mismatches 191; Indels 58; Gaps 12;

QY 6 VAVCVVRPLNSRRESLG-ETAAQVTKTDNNVIYQVDGSKSNFDRVFHGNETTNNVYEE 64
Db 10 VRVALRCRLVPKEISEGCMCLSFVPGEPQVWVGT--KSFTYDFVDPSTEQEVFNT 67
QY 65 IAAPIIDSAIQNGTIFAYGOTASGKTYTMG-----SDHLGVIPRAIHDFQKIK 117
Db 68 AVAPLKGFKGNATVLAAGTGGKTYSMGAVTASQENEPYGVIPRVQLLFXEID 127
QY 118 KPPDFEFLLRVSMEIYNETIDLLCGTQKMKPLIREDVNRNVYVADLTERVVYVTEMA 177
Db 128 KKSDFEFLKVSLEYNEEILDLCPKREKQAINREDPKGKIVGLTEKVLVADLT 187
QY 178 LKWTIKGSKRHYGKTKNQRSSRSHITFRMILSRKEGPEPNCRGSKVSHNLNVDLAG 237
Db 188 VSCLEQGNNSRTVASTAMNSQSSRSHAFIT---SLEQKKSKDNKNSFR-SKLHLVDLAG 243
QY 238 SERAAQTGAAGVRLKEGNCINRSLFILGVIKLSDGVGGFPIYVRDSEKLTILONSLGG 297
Db 244 SERQKKTAAEGDLKEGININGLLCLGNVISALGDDKGGFPAPYRDSKLTLLQDSLGG 303
QY 298 NPKTRIICTITPV--SPDETTLALQFASAKYMKNTPYVNEVSTDEALLKRYKREIMDLK 355
Db 304 NSHTLMIACVSPADSNLEETLTLRYADBARKKNKPIVN-----IDPQTAEINHLK 355
QY 356 KQLEEVSL-----ETRAQAMEKQDLAQLLEKOLLQKVONEKIENL 396
Db 356 QVQVQLQVLLQAHGGLPGSINLQSLMEKNQ--SLVEENEKLSRGLSEAAAGQT 413
```

```
QY 397 TRML-----VTSSSLTQOELKAKKRRVTWC---LGKINKMKNSNYADQFNIPNITTKTH 450
Db 414 AQMLERILITEQANERKNNAKLELRQHAACKLDLQKLVELEDOELKENVEILCNLQILIT 473
QY 451 KLSINLLREIDRSVCSDESDFNTLDTLSEIENPAT 487
Db 474 QLS-----DETVACTAAAIIDTAVBEAQVETSPET 503
```

RESULT 5

```
US-09-541-782-8
; Sequence 8, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541,782
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-541-782-8
```

Query Match 28.3%; Score 703; DB 3; Length 1066;
Best Local Similarity 34.4%; Pred. No. 1.3e-52;
Matches 187; Conservative 93; Mismatches 183; Indels 80; Gaps 17;

```
QY 6 VAVCVVRPLNSRRESL--GETAQVTKTDNNVIYQVDG--SKSNFDRVFHGNETTNNV 61
Db 20 IQVYVRPLNSRERCIRSAEVVDVVGPREVTRHLDLSKLTFFDRSGFESQKQCDV 79
QY 62 YEEIAPIIDSAIQNGTIFAYGOTASGKTYTMGSE-----DHLGVIPRAIH 110
Db 80 YSVVSPLEEVLYNGYNTVFAYGQTGTGKTHVGNETAELKSSWEDSDIGIIPRLS 139
QY 111 DIFQKIKFPDFEFLLRVSMEIYNETIDLLCGTQKMKPLIREDVNR--NVYVADLTTEE 169
Db 140 HLFDELRMW--EVEYTMRSYLELYNEELCDLSTDDTKIRIFDDSTKKGSVIIQGLEEI 198
QY 170 VVYTEMALKWTIKGSKRHYGETKQNRSSRSHITFRMILSRKEGPEPNCRGSKVYVSH 229
Db 199 PVHSKDDVYKLEKGERKERTATILMNAQSSRSHITFVIVHIRENGIEG--EDMLKIGK 256
QY 230 LNLVDLAGSRAAQTG--AAGVRLKEGNCINRSLFILGVIKLSDGVGGFPIYVRDSEKLT 288
Db 257 LNLVDLAGSENVSAGNEKGRVRETVNINQSLTLGRVITALVDR--APHVPYRESKLT 314
QY 289 RILQNSLGCNPKTRIICTITP--VSFDETLALQFASAKYMKNTPYVNEVSTDEALLKR 346
Db 315 RLQESLGGRTKTSIIATISPGKDIETSLTLEYAHRKNIQKPEVYNQKLTTKTVLKE 374
QY 347 YRKEIMDLKQOL-----EVSLETRAQAMEKQDLAQLLEB-KDLLQKVQ 389
Db 375 YTEIDKLKRDMAARDKNGIYLAEEYGETLKLSEQNREINREKMLLKLKDLQNK 434
QY 390 N-----EKIENITR---MLVTSSTLTQOELKAKKRRVTWCLGKINKMKNSNY 435
Db 435 KIFSEVSMISLEKTOELKKTTEENLLNTKGTLLTKVLTTKRY-----KEKELV 486
QY 436 ADQFNIPNITTK-----THKL--SINLREIDESVCSDESDFNTL-DTL 478
Db 487 ASHMKTEQVLTQOELIAAADLATDTHQLHGHTIERREDELDEKIRRSQDKQRMQDNL 546
QY 479 SEI 481
Db 547 EMI 549
```

GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/10/270,085
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/09/723,820
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-270-085-8

Query Match 28.3%; Score 703; DB 4; Length 1066;
Best Local Similarity 34.4%; Pred. No. 1.3e-52;
Matches 187; Conservative 93; Mismatches 193; Indels 80; Gaps 17;

QY 6 VAVCVVRPLNSREESL--GETAQVYWKTDNNVIYQVDG--SKSFNDRFVHGNFTKNV 61
DB 20 IQVTVRPLNSRERCIRSAEVDVVGPREVTVRHTLDSKLTKKFTFDRSGPESKQCDV 79
QY 62 YEEIAAIIIDSAIQVNGTIFAYGQTASGKTYTMGSE-----DHLGVIPRAIH 110
DB 80 YSVVVSPLIEVLNGYNTVFAYGQTGKTHTWVGNETAELKSSWEDSDIGIIPRALS 139
QY 111 DIFOKIKKFPDRFLLRVSYMEIYNETITDLCGTQRMKPLIREDVNR--NVYVADLTEE 169
DB 140 HLFDELMM--EVEYTMRSYLELYNEELCDLLSDTDTTKIRIFDDSTKKGSVIIQGLEEI 198
QY 170 VVYTSMALKWIITGKSRHYGETKMNORSRSHITIFRMILESEKGEPSNCEGSVKVSH 229
DB 199 PVHSKDDVYKLEKGERKRTATTLMNAQSSRSHTVFSIVVHIRENGIEG--EDMLKIGK 256
QY 230 LNLVDLAGSERRAAQTG--AAGVRLKEGNCNINRSLFILGVIKKLSGQVGGFINYRDSKLT 288
DB 257 LNLVDLAGSENVSAGNEKGIRVRETVMNINQSLTLGRVITVLDVDR--APHVPYRESKLT 314
QY 289 RILQNSLGNPKTRIICTITP--VSFDETLALQFASTAKYMKNTPPYVNEVSTDEALLKR 346
DB 315 RLLOESLGGRTKTSIIATISPGHKDIBETLSTLEYAHRAKNIQNKPEVNQKLTKKTVLKE 374
QY 347 YRKEIMDLKKQI-----EEVSLETRAQAMEKDQLAQLEE--KDLLOKVQ 389
DB 375 YTEEDLKLKRDLMARDKNGIYLAETVGEITLKLESQNRNEKMLLKALKDBLQNK 434
QY 390 N-----EKIENLTR-----MLVTSSTLTQOELKAKRRRVTCWLGKINKMKNY 435
DB 435 KIPSEVMSLVEKTOELKKEKTEENLNTKGTLLTKVLTKKRY-----KEKKEIV 486
QY 436 ADQFNIPNTITK-----THKL--SINLREIDESVCSSESVFNTL--DTL 478
DB 487 ASHMKTEQVLTQAOEILAAADLATDDTHLGHGHTIERREBELDEKIRRSQDFKDRMQDNL 546
QY 479 SEI 481
DB 547 EMI 549

RESULT 8
US-09-592-054-4
; Sequence 4, Application US/09592054
; Patent No. 6440684
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Finer, Jeffrey
; APPLICANT: Sakowicz, Roman

US-09-723-820-8
; Sequence 8, Application US/09723820
; Patent No. 6468760
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/723,820
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-723-820-8

Query Match 28.3%; Score 703; DB 4; Length 1066;
Best Local Similarity 34.4%; Pred. No. 1.3e-52;
Matches 187; Conservative 93; Mismatches 193; Indels 80; Gaps 17;

QY 6 VAVCVVRPLNSREESL--GETAQVYWKTDNNVIYQVDG--SKSFNDRFVHGNFTKNV 61
DB 20 IQVTVRPLNSRERCIRSAEVDVVGPREVTVRHTLDSKLTKKFTFDRSGPESKQCDV 79
QY 62 YEEIAAIIIDSAIQVNGTIFAYGQTASGKTYTMGSE-----DHLGVIPRAIH 110
DB 80 YSVVVSPLIEVLNGYNTVFAYGQTGKTHTWVGNETAELKSSWEDSDIGIIPRALS 139
QY 111 DIFOKIKKFPDRFLLRVSYMEIYNETITDLCGTQRMKPLIREDVNR--NVYVADLTEE 169
DB 140 HLFDELMM--EVEYTMRSYLELYNEELCDLLSDTDTTKIRIFDDSTKKGSVIIQGLEEI 198
QY 170 VVYTSMALKWIITGKSRHYGETKMNORSRSHITIFRMILESEKGEPSNCEGSVKVSH 229
DB 199 PVHSKDDVYKLEKGERKRTATTLMNAQSSRSHTVFSIVVHIRENGIEG--EDMLKIGK 256
QY 230 LNLVDLAGSERRAAQTG--AAGVRLKEGNCNINRSLFILGVIKKLSGQVGGFINYRDSKLT 288
DB 257 LNLVDLAGSENVSAGNEKGIRVRETVMNINQSLTLGRVITVLDVDR--APHVPYRESKLT 314
QY 289 RILQNSLGNPKTRIICTITP--VSFDETLALQFASTAKYMKNTPPYVNEVSTDEALLKR 346
DB 315 RLLOESLGGRTKTSIIATISPGHKDIBETLSTLEYAHRAKNIQNKPEVNQKLTKKTVLKE 374
QY 347 YRKEIMDLKKQI-----EEVSLETRAQAMEKDQLAQLEE--KDLLOKVQ 389
DB 375 YTEEDLKLKRDLMARDKNGIYLAETVGEITLKLESQNRNEKMLLKALKDBLQNK 434
QY 390 N-----EKIENLTR-----MLVTSSTLTQOELKAKRRRVTCWLGKINKMKNY 435
DB 435 KIPSEVMSLVEKTOELKKEKTEENLNTKGTLLTKVLTKKRY-----KEKKEIV 486
QY 436 ADQFNIPNTITK-----THKL--SINLREIDESVCSSESVFNTL--DTL 478
DB 487 ASHMKTEQVLTQAOEILAAADLATDDTHLGHGHTIERREBELDEKIRRSQDFKDRMQDNL 546
QY 479 SEI 481
DB 547 EMI 549

RESULT 7
US-10-270-085-8
; Sequence 8, Application US/10270085
; Patent No. 6627408

```
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6440684el motor proteins and methods for
; FILE REFERENCE: 1016
; CURRENT APPLICATION NUMBER: US/09/592,054
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Human
; US-09-592-054-4

Query Match      28.3%; Score 702; DB 4; Length 522;
Best Local Similarity 35.8%; Pred. No. 5.4e-53;
Matches 181; Conservative 83; Mismatches 187; Indels 54; Gaps 12;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNFDRVPHGNETTKNVYEE 64
DQ 30 VRVALRCRPLVPKEISEGCMCLSFVPGEPVVVGTD--KSTYDFVDFPSTEQEEVFNT 87
QY 65 IAAPIDSAIOGYNGTIFAYGOTASGKTYVMG-----SEDLGVIPRAIHDFQKIK 117
DQ 88 AVAPLIKGVFKGNATVLAAYGOTGSKTYSMGAYTAEQENEPVTVGVIPRVQLLFKEID 147
QY 118 KPDPREFLRVSMETYNETITDLCGTOCKMKPLIREDVNRNVVADLTTEEVVYSEMA 177
DQ 148 KKSDFEFTLVKSYLEIYNEEILDLCPREKAKINIREDPKGIKIVGLTEKTVLVALDT 207
QY 178 LKWIITGKSRHYGTCKMORSRSHITPRMILESEKGEPCNCGSVKVSHLNLDVLAG 237
DQ 208 VSCLEQGNNSRTVASTAMNSQSSRSHAITISLEQKKS-----KNSFSRKLHLVDLAG 263
QY 238 SRAAQTGAAGVRLKEGNCINRSLFILGOVKKLSDGQVGGFVINYRDSKLTILQNSLGG 297
DQ 264 SERQKTKAEGRLKEGININRGLLGNVISALGDDKGGFVPYRDSKLTLLQDSLGG 323
QY 298 NPKTRIICTITPV--SPDETALOFASTAKYMKNTPVYNEVSTDEALLKRYRKEIMDLK 355
DQ 324 NSHTLMIACVSPADSNLEETLWLRADRKIKKPIVN-----IDPOTAEINHLK 375
QY 356 KQLEEVSL-----ETRAQMEKQDLAQLLEKDLQKVNQKNIENL 396
DQ 376 QVQVQLQVLLQAHGGLPGSTVPESENLOSLMEKNQ--SLVEENEKLSRGLSEAAGOT 433
QY 397 TRLM---VTSSSLTLQQLKAKRKRRTVCLGKINK---MKNVYADQFNPTITTK- 448
DQ 434 AQMLERIIWTEQANEKMAKLEELRQHAACKLDLQKLVTLEDEQELKENVEIICNLQQLI 493
QY 449 THKLSINLLREIDESVCSSESVFSN 473
DQ 494 TQKLA--ALEGTEQKLISEEDLIEH 517

RESULT 9
US-09-592-054-2
; Sequence 2, Application US/09592054
; Patent No. 6440684
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Finer, Jeffrey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6440684el motor proteins and methods for
; FILE REFERENCE: 1016
; CURRENT APPLICATION NUMBER: US/09/592,054
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1232

; TYPE: PRT
; ORGANISM: Human
; US-09-592-054-6

Query Match      28.0%; Score 695; DB 4; Length 473;
Best Local Similarity 39.7%; Pred. No. 1.9e-52;
Matches 172; Conservative 70; Mismatches 155; Indels 36; Gaps 10;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNFDRVPHGNETTKNVYEE 64
DQ 10 VRVALRCRPLVPKEISEGCMCLSFVPGEPVVVGTD--KSTYDFVDFPSTEQEEVFNT 67
QY 65 IAAPIDSAIOGYNGTIFAYGOTASGKTYVMG-----SEDLGVIPRAIHDFQKIK 117
DQ 68 AVAPLIKGVFKGNATVLAAYGOTGSKTYSMGAYTAEQENEPVTVGVIPRVQLLFKEID 127
QY 118 KPDPREFLRVSMETYNETITDLCGTOCKMKPLIREDVNRNVVADLTTEEVVYSEMA 177
DQ 128 KKSDFEFTLVKSYLEIYNEEILDLCPREKAKINIREDPKGIKIVGLTEKTVLVALDT 187
QY 178 LKWIITGKSRHYGTCKMORSRSHITPRMILESEKGEPCNCGSVKVSHLNLDVLAG 237
DQ 188 VSCLEQGNNSRTVASTAMNSQSSRSHAITISLEQKKS-----KNSFSRKLHLVDLAG 243
QY 238 SRAAQTGAAGVRLKEGNCINRSLFILGOVKKLSDGQVGGFVINYRDSKLTILQNSLGG 297
DQ 244 SERQKTKAEGRLKEGININRGLLGNVISALGDDKGGFVPYRDSKLTLLQDSLGG 303
QY 298 NPKTRIICTITPV--SPDETALOFASTAKYMKNTPVYNEVSTDEALLKRYRKEIMDLK 355
DQ 304 NSHTLMIACVSPADSNLEETLWLRADRKIKKPIVN-----IDPOTAEINHLK 355
QY 356 KQLEEVSL-----ETRAQMEKQDLAQLLEKDLQKVNQKNIENL 396
DQ 356 QVQVQLQVLLQAHGGLPGSTVPESENLOSLMEKNQ--SLVEENEKLSRGLSEAAGOT 413
QY 397 TRLM---VTSSSLTLQQLKAKRKRRTVCLGKINK---MKNVYADQFNPTITTK- 449
DQ 414 AQMLERIIWTEQANEKMAKLEELRQHAACKLDLQKLVTLEDEQELKENVEIICNLQQLI 473
QY 450 HKLSINLLREIDESVCSSESVFSN 487
DQ 474 TQLS-----DETACMAALDAVEQEAQVETSPET 504

RESULT 10
US-09-592-054-6
; Sequence 6, Application US/09592054
; Patent No. 6440684
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Finer, Jeffrey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6440684el motor proteins and methods for
; FILE REFERENCE: 1016
; CURRENT APPLICATION NUMBER: US/09/592,054
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Human
; US-09-592-054-6
```

Db 8 VRVALRCPLVPKXEISEGCOMCLSFVPEGPQVVVGTD--KSFTYDFVDFSTBQEEVFNT 65
 Qy 65 IAPIIDSAIQYNGTIFAYGOTASGKTYTMMG-----SEDLGVIPRAIHDFQIK 117
 Db 66 AVAPLKGVPKYNATVLAGYQGTSGKTYSMGAYTAQENEPVGVIPVIOQLLFEID 125
 Qy 118 KFPDRPFLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNRYVADLTVEEVYVTSEMA 177
 Db 126 KKSDFETLVSVLEIYNEEILDLCPSREKAQINREDPKGKIVGLTEKTVALDIT 185
 Qy 178 LKMITKGEKSRHYGETKQNRSSRSHYTFMILESREKGPSCGKGVKSHLNLVDLAG 237
 Db 186 VSCLEQGNNSRTVASTAMNSQSSRSHAFITISLEQRKSD-----KNSFSRSLHLVDLAG 241
 Qy 238 SERAAQTGAAGVRLKEGCNINRSIFILGQVVKLSGQVGFNYRDSKLTIRLONSLGG 297
 Db 242 SEROKTKAAGDRLEKGININRGLLCLGNVISALGDDKGGFVYRDSKLTIRLQDSLGG 301
 Qy 298 NPXTRIICTIPV--SPDETALQFASTAKYMKNTPYVNEVSTDEALLKRYRKBMIDLK 355
 Db 302 NSHTLMIACVSPADSNLEETLNTLRVADRARKINKPKIVN-IDPQTAEHLNHLKQOV--- 356
 Qy 356 KQLEEVSVLETRAQAM-----EKDQLAQLLEBKDLLOKVONEKIENITRMLVTSSSLTLQ 409
 Db 357 QQLQVLLQLAHGGTLPGSITVPSENIQSLMEKNQSLVEBNEK---LSRGLSEAAQT-- 411
 Qy 410 QELKAKRKRVTW 422
 Db 412 ----AQMLEKIIW 420

RESULT 11

US-09-914-259-24

; Sequence 24, Application US/09914259

; Patent No. 6495336

; GENERAL INFORMATION:

; APPLICANT: Makowski, Lee

; APPLICANT: Hyman, Paul

; APPLICANT: Williams, Mark

; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES

; FILE REFERENCE: 8471-010-999

; CURRENT APPLICATION NUMBER: US/09/914,259

; CURRENT FILING DATE: 2000-11-21

; NUMBER OF SEQ ID NOS: 180

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 24

; LENGTH: 1031

; TYPE: PRT

; ORGANISM: Strongylocentrotus purpuratus

US-09-914-259-24

Query Match

Best Local Similarity 26.8%; Score 664.5; DB 4; Length 1031;

Matches 192; Conservative 80; Mismatches 164; Indels 161; Gaps 18;

Qy 1 AEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVYQVDSKSFNDRVPHGNETTKN 60
 Db 5 AECNIKVV-CVRPMNATEQN--TSHICTKFISEEQVQI-GGKLNMFDRIFKENTTQEE 59
 Qy 61 VYEBIAPIIDSAIQYNGTIFAYGOTASGKTYTM---MGSEDLGVIPRAIHDFQIK 117
 Db 60 VYNKAARQIVKDVLDGNGTIFAYGQTSNGTFTMEGVMPQYMGIIIPRVQDFINHY 119
 Qy 118 KFPDR-BPFLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNRYVADLTVEEVYVTSEM 176
 Db 120 QMDESLEFHIKVSFEIYMDIRDL--DVSKTNLSVHEDKRVFPVKGATERPASSPEE 177
 Qy 177 ALKMITKGEKSRHYGETKQNRSSRSHYTFMILESREKGPSCGKGVKSHLNLVDL 235
 Db 178 VMDVTEEGKSRNHIAVTNNHSSRSHSIF--LIQVKOE---NMETKKLSGKLVLVDL 231
 Qy 236 AGSERAAQTGAAGVRLKEGCNINRSIFILGQVVKLSGQVGFNYRDSKLTIRLONSL 295

Db 232 AGSEKVKTAGETVLDDEAKNINKSLSALGNVISALADGK-KSHIPYRDSKMTIRILOESL 290
 Qy 296 GGNPKTRIICTIPVSPFD--ETILTALQFASTAKYMKNTPYVNEVSTDEALLKRYRKE--- 350
 Db 291 GGNARTIVICCSFSPSESESKSTLMFGQRAKTIKNTVTVNMELTAEERWRYEKEKEK 350
 Qy 351 -----IMDLKKQL-----LAQLLEBKDLLOKQV-----EE 360
 Db 351 NGRSLKAQLLILELORWRAGESVPVKEQGNKNDILKEMMKPKQMTVHVSEBEKNKWE 410
 Qy 361 VSLLETRAQAMEKQD-----LAQLLEBKDLLOKQV----- 389
 Db 411 EKVKLYEQDLDEKSEIDNQSRLEKTKQKQMLESEILLSMQRDYELLOSQMGRLEAENAA 470
 Qy 390 -----NEKI-----ENLTRMLVTSSSLTLQOEL 412
 Db 471 AKEAEKEVLOALEEMAVNYDEKSKVEDKRNMMNETLSEEVNEKMTALHTTSTTELQKLOEL 530
 Qy 413 KAKRKRRTWC-----LGKINKMKNNSYADQNFPTNITTKTKLSINLLREIDE 462
 Db 531 ECHRRRRIITEMASLLKDLGEIGTALGGNAADM-----KPNVENIEKVD 575

RESULT 12

US-09-914-259-19

; Sequence 19, Application US/09914259

; Patent No. 6495336

; GENERAL INFORMATION:

; APPLICANT: Makowski, Lee

; APPLICANT: Hyman, Paul

; APPLICANT: Williams, Mark

; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES

; FILE REFERENCE: 8471-010-999

; CURRENT APPLICATION NUMBER: US/09/914,259

; CURRENT FILING DATE: 2000-11-21

; NUMBER OF SEQ ID NOS: 180

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 19

; LENGTH: 975

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

US-09-914-259-19

Query Match

Best Local Similarity 26.6%; Score 660.5; DB 4; Length 975;

Matches 193; Conservative 76; Mismatches 200; Indels 55; Gaps 20;

Qy 1 AEGAVAVCVVRPLNSREESLGETAQVYWKTDNNV--IYQVDGSKSFNDRVPHGNET 57
 Db 9 AEDSIKVV-CRFRPLNDESEKAGSKFV--KFPNNVEENCISIAS-KVYLEDKVFKNAS 64
 Qy 58 TKVYEBIAPIIDSAIQYNGTIFAYGOTASGKTYTM---MGSEDLGVIPRAIHDFQ 114
 Db 65 QEKVYNEAASIVTDVLAGNGTIFAYGQTSNGTFTMEGVDSKQVGIIPRVINDIFN 124
 Qy 115 KIKFPP-DRBPFLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNRYVADLTVEEVY 173
 Db 125 HIYAMEVNLPHIKVSYEYIMDKIRDL--DVSKVN--LSVHEDKRVYPVKGATERFVSS 182
 Qy 174 SEMALKMITKGEKSRHYGETKQNRSSRSHYTFMILESREKGPSCGKGVKSHLNL 232
 Db 183 PEDVFEIEEGKSRNHIAVTNNHSSRSHSIFLNKQ-----ENLECKLSGKLVL 236
 Qy 233 VDLAGSERAAQTGAAGVRLKEGCNINRSIFILGQVVKLSGQVGFNYRDSKLTIRLQ 292
 Db 237 VDLAGSEKVKTAGETVLDDEAKNINKSLSALGNVISALADGN-KTHIPYRDSKLTIRLQ 295
 Qy 293 NSLGNPKTRIICTIPVSPFD--ETILTALQFASTAKYMKNTPYVNEVSTDEALLKRYRKE 350
 Db 296 ESLGNGARTTIVICCSFSPSESESKSTLMFGQRAKTIKNTVTVNMELTAEERWRYEKEKE 355
 Qy 351 ---IMDLKKQLEEVSVLE--TRAQAMEKQDLAQLLEBKDLLO-KVQNEKIENITRMLVTS 405

Db 356 KKNARLKGKVEKLEIELARWRAGETVKAEEQINMEDLMESTPNLEVE-AAQTAAAEAA 414
QY 406 LTLOQELKA-----KXKRVVWCLGKINKM-----KNSNYADQF-----N 440
Db 415 LAAQRTALANASAVVNEQARLATECERLYOOLDDKDBEINQOQYABQLKEQVMEQBE 474
QY 441 IPTNITTKHLSINLLREIDESVCSSESVFNSNTLDTLSEIENW 484
Db 475 LIANARREYETLOSEWARIQENESAKEEV-KEVLQALEELAVN 517

RESULT 13

US-09-572-191-2
; Sequence 2, Application US/09572191
; Patent No. 6355466
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6355466el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/572,191
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
US-09-572-191-2

Query Match 26.4%; Score 656; DB 4; Length 1388;
Best Local Similarity 37.8%; Pred. No. 2.4e-48;
Matches 201; Conservative 69; Mismatches 168; Indels 94; Gaps 20;

QY 2 BEGAVAVCVVRVPLNSREESL-GETAQVYWKTDNNVIYQVDGS-----KSFNFD 49
Db 23 EGDALKVFRIRPPAERSGSADGE-----QNLCLSVLSSTSLRLHSNPEPTFTFD 73
QY 50 RVFHGNETTKVYIEAIPIDSAIQYNGTIFAYGQTASGKTYTMMG-----SEHIL 102
Db 74 HVADVDTTQESVFATVAKSIVESCMGNGTIFAYGQTSGKTFMTMGPSDFNSHNL 133
QY 103 GVIPRAIHDF-----QKIKFPDRELLRVSYMEIYNETITDLCGQTKMKPLIREDV 157
Db 134 GVIPRSFYLSLDREKEKAGAGKSLCKSCSFIEYNEQIYDIL--DSASAGLYLREHI 191
QY 158 NRVVYVADLTBEVVYVTSMAKWKITGKSRHYGETKMNQSSRSHTIFRMIRESKGE 217
Db 192 KKGVFVVGAVQVTSAAEAQVLSGGWRNRVASTSMNRESSRSHAVFTTIESMEK-- 249
QY 218 PSNCEGSVKVSHNLVDLAGSRAAQTGAAGVRLKEGNNINRSLFILQGVTKKLSGDQVG 277
Db 250 -SNEIVNRTSLNLVDLAGSERQKDTAEGMRLKEAGNINRSLSCLGQVITALVD--VG 306
QY 278 G----FINYRDSKLTILQNSLGNPKTRIICTITPVS--FDETLTALQFASTAKYMKNT 331
Db 307 NGKQHVYCYRDSKLTFLRLDSLGGNAKTAIIANVHPGRCFGETLSTLNFQAKLIKNK 366
QY 332 PYVNE-----VSTDEALKRYRKEIMDLKK-QLEEVSLETR-----AQAMEKDQLAQL-- 379
Db 367 AVNEDTQGNVSQLAQEVKRLKEQLAELASGQTPPEPFLTRDKKNTYMEYFQEAFLFFK 426
QY 380 ----BEKOLLQVQNEKIENLRMLVTSSTLTQOELKAKKRRVTCWLGKINKMKNSNY 435
Db 427 KSEQEKSLIEKV-----TQLEDLTLLKKEKFT-----QSNKMIVKFR 463
QY 436 ADQFNIPNITTKTKHLSI-NLLREIDESVCSSESVFNSNTLDTLSE-LEWNP 485
Db 464 EDQI-----IRLEKHKESRGGLPEEQDRLISE---LRNEIQTUREQIEHPH 508

RESULT 15

US-09-723-219-2
; Sequence 2, Application US/09723219
; Patent No. 6391613
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6391613el motor proteins and methods for

RESULT 14
US-09-723-262-2
; Sequence 2, Application US/09723262
; Patent No. 6379912
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6379912el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/723,262
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
US-09-723-262-2

Query Match 26.4%; Score 656; DB 4; Length 1388;
Best Local Similarity 37.8%; Pred. No. 2.4e-48;
Matches 201; Conservative 69; Mismatches 168; Indels 94; Gaps 20;

QY 2 BEGAVAVCVVRVPLNSREESL-GETAQVYWKTDNNVIYQVDGS-----KSFNFD 49
Db 23 EGDALKVFRIRPPAERSGSADGE-----QNLCLSVLSSTSLRLHSNPEPTFTFD 73
QY 50 RVFHGNETTKVYIEAIPIDSAIQYNGTIFAYGQTASGKTYTMMG-----SEHIL 102
Db 74 HVADVDTTQESVFATVAKSIVESCMGNGTIFAYGQTSGKTFMTMGPSDFNSHNL 133
QY 103 GVIPRAIHDF-----QKIKFPDRELLRVSYMEIYNETITDLCGQTKMKPLIREDV 157
Db 134 GVIPRSFYLSLDREKEKAGAGKSLCKSCSFIEYNEQIYDIL--DSASAGLYLREHI 191
QY 158 NRVVYVADLTBEVVYVTSMAKWKITGKSRHYGETKMNQSSRSHTIFRMIRESKGE 217
Db 192 KKGVFVVGAVQVTSAAEAQVLSGGWRNRVASTSMNRESSRSHAVFTTIESMEK-- 249
QY 218 PSNCEGSVKVSHNLVDLAGSRAAQTGAAGVRLKEGNNINRSLFILQGVTKKLSGDQVG 277
Db 250 -SNEIVNRTSLNLVDLAGSERQKDTAEGMRLKEAGNINRSLSCLGQVITALVD--VG 306
QY 278 G----FINYRDSKLTILQNSLGNPKTRIICTITPVS--FDETLTALQFASTAKYMKNT 331
Db 307 NGKQHVYCYRDSKLTFLRLDSLGGNAKTAIIANVHPGRCFGETLSTLNFQAKLIKNK 366
QY 332 PYVNE-----VSTDEALKRYRKEIMDLKK-QLEEVSLETR-----AQAMEKDQLAQL-- 379
Db 367 AVNEDTQGNVSQLAQEVKRLKEQLAELASGQTPPEPFLTRDKKNTYMEYFQEAFLFFK 426
QY 380 ----BEKOLLQVQNEKIENLRMLVTSSTLTQOELKAKKRRVTCWLGKINKMKNSNY 435
Db 427 KSEQEKSLIEKV-----TQLEDLTLLKKEKFT-----QSNKMIVKFR 463
QY 436 ADQFNIPNITTKTKHLSI-NLLREIDESVCSSESVFNSNTLDTLSE-LEWNP 485
Db 464 EDQI-----IRLEKHKESRGGLPEEQDRLISE---LRNEIQTUREQIEHPH 508

```
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/723,219
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
; US-09-723-219-2

Query Match      26.4%; Score 656; DB 4; Length 1388;
Best Local Similarity 37.8%; Pred. No. 2.4e-48;
Matches 201; Conservative 69; Mismatches 168; Indels 94; Gaps 20;

QY      2  EGA VAVCVVRRLNREESL-GETAQVYMKTDNNVIYQVDGS-----KGFNFD 49
Db      23  EGD AIVFVRIRPPAERSGADGE-----QNLCLSVLSSTSLRLHSNPEPKPTFD 73

QY      50  RVFHGNETTKNVVEETAAPIDSAIQYNGTIFAYGQTASGKTYTMWG-----SEDHL 102
Db      74  HVADVDVTQBSVFATVAKSIVESCMSGYNGTIFAYGQTGSGKTFMMPSESNFHNLR 133

QY      103 GVIPRAIHDIF-----QKIKFPDPREPLRVSYMEIYNETITDLCGTQKMKPLIREDV 157
Db      134 GVIPRFEYLFSLIDREKEKAGAKSPFLCKSPFIEIYNEQIYDLL--DSASAGLYLREHI 191

QY      158 NRNVYVADLTBEVVYVYTSMAKWTITGKGRHYGETKQNRQSRSHYTFRMILESRKGE 217
Db      192 KGVFVVGAVEQVVTSAEAYQVLSGGWRNRVASTSMNRESSRSHAVFTITTESMEK-- 249

QY      218 PSNCEGSVKVSHLNLVDLAGSERAQGTGAAGVRLKEGCNINRSLFILGOVIKLSGQVG 277
Db      250 -SNEIVNIRTSLNLVDLAGSERQKQTHAEGMRLEKAGNINRSLSCIGQVITALVD--VG 306

QY      278 G-----FINYRDSKLTRILQNSLGNPKTRIITTPVS--FDETTLTALQFPAKYMKNT 331
Db      307 NGQRHVCYRDSKLTFLRDSLGNAKTAAIANVHPSRCFGETLSTLNFQAQAKLIKX 366

QY      332 PYVNE-----VSTDEALLKRYRKEIMOLKK-QLEEVSLETR---AQMEXDQLAQLL-- 379
Db      367 AVVNEDTQGNVSQLQAEVKRLKEQLAELASGQTPPESTFLTRDKKKKNYMEYFOEAMLFFK 426

QY      380 ----EKKDLQKQWQEKIENLTRMLVTSSSLTLQOELKAKRGRVTWCLGKINKMKNSNY 435
Db      427 KSEQEKSLIEKV-----TQLEDLTLKKEFI-----QSNRMIVKFR 463

QY      436 ADQFNIPNTITTKTKHLSI-NLLREIDESVCSDFSNITDITLSE-IEWNP 485
Db      464 EDQI-----IRLEKLHESRGGLFEEQDRLLSE---LRNEIQTLREQIEHHP 508
```

Search completed: July 29, 2004, 09:42:49
Job time : 10.6277 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:39:47 ; Search time 24.324 Seconds
(without alignments)
6280.361 Million cell updates/sec

Title: US-10-045-631b-88_COPY_2_488

Perfect score: 2483

Sequence: 1 AEEGAVAVCVVRPLNSREE.....SDVFNLTDLTSLIEWNPAT 487

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp:*
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pcp:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp:*
8: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pcp:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pcp:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pcp:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp:*
12: /cgn2_6/ptodata/2/pubpaa/US09C_NEW_PUB.pcp:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp:*
16: /cgn2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pcp:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	973.5	39.2	1382	16	US-10-437-963-176714
2	878	35.4	694	12	Sequence 176714, A
3	750.5	30.2	1006	16	Sequence 59725, A
4	718.5	28.9	420	16	Sequence 147933, A
5	717.5	28.9	1237	15	Sequence 114373, A
6	714.5	28.8	955	12	Sequence 33, Appl
7	713.5	28.7	1232	14	Sequence 62672, A
8	713.5	28.7	1232	16	Sequence 670, App
9	713.5	28.7	1235	15	Sequence 2153, App
10	710.5	28.6	1029	16	Sequence 8, Appl
11	702.5	28.3	1232	14	Sequence 182113, A
12	702.5	28.3	1232	14	Sequence 2, Appl
13	702.5	28.3	1232	14	Sequence 664, App
14	691.5	27.8	965	16	Sequence 669, App
15	689	27.7	672	16	Sequence 169150, A
					Sequence 1664, App

16	671.5	27.0	966	12	US-10-425-114-62697
17	664.5	26.8	1031	14	Sequence 62697, A
18	664.5	26.8	1031	15	Sequence 24, Appl
19	663.5	26.7	406	15	Sequence 113, App
20	660.5	26.6	517	15	Sequence 3789, App
21	660.5	26.6	975	14	Sequence 12881, A
22	660.5	26.6	975	15	Sequence 19, Appl
23	658	26.5	757	15	Sequence 108, App
24	658	26.5	757	12	Sequence 366, App
25	656	26.4	1388	14	Sequence 316, App
26	656	26.4	1388	12	Sequence 2, Appl
27	656	26.4	1388	15	Sequence 82, Appl
28	656	26.4	1388	16	Sequence 164, App
29	654.5	26.4	963	14	Sequence 20, Appl
30	654.5	26.4	963	15	Sequence 109, App
31	653.5	26.3	963	14	Sequence 22, Appl
32	653.5	26.3	963	15	Sequence 111, App
33	652.5	26.3	821	16	Sequence 114370, A
34	650	26.3	1362	12	Sequence 2, Appl
35	650	26.2	1826	16	Sequence 1477, App
36	648.5	26.1	1022	16	Sequence 133108, A
37	643.5	25.9	2552	16	Sequence 129734, A
38	642.5	25.9	967	14	Sequence 21, Appl
39	642.5	25.9	967	15	Sequence 110, App
40	640	25.8	957	14	Sequence 16, Appl
41	640	25.8	957	15	Sequence 105, App
42	638.5	25.7	1011	12	Sequence 24, Appl
43	638.5	25.7	1011	12	Sequence 58, Appl
44	638.5	25.7	1032	14	Sequence 26, Appl
45	638.5	25.7	1032	15	Sequence 115, App

ALIGNMENTS

RESULT 1

US-10-437-963-176714
; Sequence 176714, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 176714
; LENGTH: 1382
; TYPE: PPT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_74437C.1.pep
US-10-437-963-176714

Query Match	39.2%	Score	973.5	DB	16	Length	1382
Best Local Similarity	46.7%	Fred. No.	4e-74				
Matches	232	Conservative	74	Mismatches	140	Indels	51
Gaps	15						
QY	6	VAVCVVRPLNSRESLCTAQVYWKTNVNI-YQVDGSKGFNDFRVPHGNETTKNVYEE	64				
Db	4	IHVAVRPLTA--EDAGSSP---WRVSGNAIALSTQPSIRFDFRIFEGECRTADVGA	58				
QY	65	IAAIIIDSAIGVNGTIPAYGQTASGKTYTMWGSDEHLGVIPRAIHDFQIKKPPDEF	124				
Db	59	RTKHIVDSAVRGFTGVAYGQTNSTGKTYTMWGSDEHLGVIPRAIHDFQIKKPPDEF	124				

125	LLRVSYMEIYNETTDL	CGTQKMKPLI	IREDVNNTVADLT	TEVVYTSEMA	KWITKG	184
119	LLRVSYMEIYNEE	INDLL--	VPEHRKQIHESI	ERGIYVAGLRE	IVTCEPQVLE	176
185	EKSRHIYGETKMNQ	SSRSHTT	IFRMLSRKGRS	---NCEGS	VKVSHNLNVD	140
177	ESHRHIGETNNV	SSRSHTT	IFRNVIESREK	VDSEAGESC	D--AVRVSNL	140
241	AAOTGAAGVRL	KEGCNINRS	FIILGQVTKKLS	SDG--QVGG	FINYRDSKLT	198
236	AAKTGAEGVRL	KEGSHINKSL	MTLGTIVIKL	SEGIEGQGHV	PYRDSKLT	198
299	PKTRIICIT	--PVSF	DETLTALOPAST	AKYMKNTPY	NEVSTDBALL	198
296	ANTAII	CNITLAQ	HADETKSSLO	FASRALR	VTNCACVNE	198
357	QL-----	EEVSL	ETRAQAMEKD	QALLEBK	DLLOKQNEKI	198
356	KIRSELEK	ERISLE	-----	LEEEK	KAKGQOR	198
411	E-----	LKAKR	KERVTC	LGKINK	MKNNSYADQ	198
402	ERDR	TTVSSKN	KRLTWC	PGLLSR	QFGQVLES	198
462	ESVCS	ES-----	DVFSN	473		
462	OES	CES	SIKH	VT	DAYSS	478

RESULT 2

```

US-10-425-114-59725
; Sequence 59725, Application US/10425114
; Publication No. US20040034889A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ. ID NOS: 73128
; SEQ. ID NO 59725
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3912-011-C4_FLI.pep
US-10-425-114-59725

```

Db	208	EQVFELLQLEA	NPHFGETNMNMRSSSHIFRMVIESSGKDQDQDGD-ALRVSVLNLVD	266
Qy	235	LAGSERAATGA	AGVRLKEGCINNRSPILGQVIKKLSD-QVGGFNYRUSKULTRIILOQ	293
Db	267	LASERIIKTGA	EVRLENGKYINKSLMILGNVINKLSENGKQRGHIPYRDSKULTRIQQ	326
Qy	294	SIGGNPKTRICI	TTP--VSFDEPRTALOFASTAKYMKNTPYNVEVSTDEALLKRYREI	351
Db	327	ALGNNAKTSIIC	TAAPEEIHETRTGTLOFSRAKVCNSCAQVNEILTDAALLKRQKLEI	386
Qy	352	MDLKQO-----	LEEVSLETR----AQAMEKDQLACLBLEKDLLOKVQVNEKTENLTRM	399
Db	387	ELARKKLGSGH	SEGLEQVVLKLRNDMHKSLELERDLAMELEDERKLRL-----	433
Qy	400	LVTSSSLTLQOE	LKA 414	
Db	434	MTLEQHLLTEQO	KLEFA 448	

RESULT 3

```

US-10-437-963-147933
; Sequence 147933, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 147933
; LENGTH: 1006
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(1006)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_48414C.1.pep
US-10-437-963-147933

```

Query Match	30.2%	Score	750.5	DB	16	Length	1006
Best Local Similarity	40.8%	Pred. No.	3.7e-55				
Matches	182	Conservative	86	Mismatches	139	Indels	39
Gaps							
Qy	1	AEEGAVACVVRPLNGREESLGETAQVWKMTDNNVIYQVDSK-SFNFDRVFHGNETTK	59				
Db	64	AAKESVAVAFRPLSPREVRGE-KIAWADGETVARSEQSNLAYAYDRVFGPTTTR	121				
Qy	60	NVTEIIAAPIDSIAIQVNGTIFAYGQTASCTYTMGSEDLGVIPRAIHDFOKIKF	119				
Db	122	HIYDAVQYVVGAMKGINGTIFAYGVTSSKTHMGDQSPGVIPLAVKDIFNLIQBT	181				
Qy	120	PDREFLLRVSMETNYETITDLCGTQMKPLIIREDDVNRNVYVADLFEVYVTSEMALK	179				
Db	182	PNREFLLRVSLIYNEVNDLL-NPAGONLRIREDLQGTI-VBGIKEAVLSPVHALS	238				
Qy	180	WTKGE--KSRHYGETMNRQSRSHTFIMLLESREKGEPSCEGSKVSHLNLVDLIA	236				
Db	239	LIAAGEVFSELRHVGSFNFLNLSRSRSHTFITLTIESSPRGQSEAE-AVTLSQLNLIDIA	297				
Qy	237	GSERAAQTGAAGVRLKEGCNINRSLIFILGVIKKLSDQGVGFINYRDSKLTFRILQNSLIG	296				
Db	298	GSR--SSRVETAGVHKEGYSINKSLTLGVKISLTD-EKATHIFPRDSKLTFRILKSSLS	355				

; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 62672
; LENGTH: 955
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3689-241-All_FLI.pcp
US-10-425-114-62672

Query Match 28.8%; Score 714.5; DB 12; Length 955;
Best Local Similarity 39.6%; Pred. No. 4.3e-52;
Matches 193; Conservative 69; Mismatches 162; Indels 63; Gaps 15;
QY 1 AEEGAVACVVRPLNSRBSLGETAQVYWK--TDNNVIYQVDGS-----KSNFDRVPH 53
DB 33 AKBEKFTVVRPLSKKE--LAAKDVAWECADTQTLTKGPAQDRAAPMSFTDFKVF 90
QY 54 GNETTKNVEEIAAPIIDSIAIQYNGTIFAYGQTASGTYTMMGSEDLGLVPIRAIHDI 113
DB 91 PACQTDVVEEGAKDVAMSALTGINATIFAYGQTSSGKTFMR-----GVTSAVSDIY 144
QY 114 QTKKFPDRFLRVSYMEIYNETITDLGCTQKMKPLIREDVNRNYYVADLTEEYVYT 173
DB 145 RHIDSTPEREFVTKISAMEIYNETIVKDLL--RPDSAPFLRLDDPEKGTIVEKQEEAKD 202
QY 174 SEMALKWTKGKSRHYGETKMNQSRSHITFRMILESRKGPSPNCEGSKVSHNLV 233
DB 203 SQHLRHUISCEBQVQVGETALNDTSRSHQILKTLLESRLR-EVSCVKSF-VANLNFV 260
QY 234 DLGSEAAQTGAAGVRLKEGNCINRSILFQGVIKKLSQGVGFYNYRDSKLTILQNL 293
DB 261 DLGSEAAQTHAIGARLKEGCHINRSLLTITVIRKLSGKSGHIPYRDSKLTILQL 320
QY 294 SLGNPKTRICTITP--VSFDETLTALQPASTAKYMNTPYNEVSTDEALLKRYKEI 351
DB 321 SLGNMARTAICTMSPALTHVEQSRNTLFFATCAKEVNTAKVNNVSDKQLVKHLOT 380
QY 352 MDLKKQLEEVSLERAQ--AMEKDQALQLE-EXDLQKQVONEKIENLTRLMTVSSSLT 408
DB 381 ARLESELRTPDRGSSDILMEKDKIRQMEIIEELRKQKD-----NA 424
QY 409 QOELKAKRRVTVT---W-----CL---GKI-----NKMKNYADQNPITNI 445
DB 425 QSELEELRKKVDHQPGWNPFDSPQARKKCLTFSGSLEPSNKNKNMSSIRQSSTAPML 484
QY 446 TTKTHKL 452
DB 485 KHEIRKL 491

RESULT 7
US-10-116-712-670
; Sequence 670, Application US/10116712
; Publication No. US20030194764A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Switzer, Ann
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.568
; CURRENT APPLICATION NUMBER: US/10/116,712
; CURRENT FILING DATE: 2002-04-07
; NUMBER OF SEQ ID NOS: 670
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 670
; LENGTH: 1232

; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-116-712-670
Query Match 28.7%; Score 713.5; DB 14; Length 1232;
Best Local Similarity 35.1%; Pred. No. 7.7e-52;
Matches 182; Conservative 86; Mismatches 191; Indels 59; Gaps 11;
QY 6 VAVCVVRPLNGREESLG-ETAQVYWKTDNNVIYQVDGSKSNFDRVFHGNETKNVYEE 64
DB 10 VRVALRCRPLVPKEISEGQCMCLSFVPGEPQVVGTD--KSFTYDFVDFPSTQEVEFT 67
QY 65 IAAPIIDSIAIQYNGTIFAYGQTASGTYTMMG-----SEDLGLVPIRAIHDIQKIK 117
DB 68 AVAPLIKGVFKGNATVLAAYGQTSGKTYSMGGYTAEOENETVGVIPRVIQILLKREID 127
QY 118 KFPDRFLRVSYMEIYNETITDLGCTQKMKPLIREDVNRNYYVADLTEEYVYSEMA 177
DB 128 KXSDFTFLKVSLEIYNEIILLDLCPSEKAGINIREDPKEGKIVGLTEKTVLVALDT 187
QY 178 LKWITTKGKSRHYGETKMNQSRSHITFRMILESRKGPSPNCEGSKVSHNLVLDLAG 237
DB 188 VSCLEQGNNSRTVASTAMNSQSSRSRSHAITFISLEQKKSD---KNSFSRSLHLVDLAG 243
QY 238 SERAAQTGAAGVRLKEGNCINRSILFQGVIKKLSQGVGFYNYRDSKLTILQNL 297
DB 244 SERQKTKAEGDLKEGININRGLLCLGNVISALGDDKGGFVYRDKLTRLQDSLGG 303
QY 298 NPXTRICTITP--SFDETLTALQPASTAKYMNTPYNEVSTDEALLKRYKEIMDLK 355
DB 304 NSHTLMACVSPADSNLEETINTLRVADRARKIKNPIVN-----IDPQTAELNHLK 355
QY 356 KQLEEVSL-----ETRAQAMEKDQALQLEEKDOLLKQVONEKIENL 396
DB 356 QVQVQQLQVLLQAHGGTLPGSITVPESENLSLMKNQ--SLVEENKLSRGLSEAAQGT 413
QY 397 TRML-----VTSSSITLQELKAKRRVTVCLGKINK---MKSNYADQNPITNITKT 449
DB 414 AQMLRIILTEQANEKMNKALELRQHAACKLDLQKLVETLEQDELKENVEIICNLOOLI 473
QY 450 HKLSINLLREIDRSVCSSESVFSNTLDTLSEIENPAT 487
DB 474 TQLS-----DETVMCAAIDTAVQEQAQVETSPET 504

RESULT 8
US-10-408-765A-2153
; Sequence 2153, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2153
; LENGTH: 1232
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2153

Query Match 28.7%; Score 713.5; DB 16; Length 1232;
Best Local Similarity 35.1%; Pred. No. 7.7e-52;
Matches 182; Conservative 86; Mismatches 191; Indels 59; Gaps 11;

Db 143 HIENPTERDPIIKISAMEIYNEIVKOLL--RPESTNLRLLDDPEKGTIVKLEEBEIAKDS 200
 Qy 175 EMALKWITKEKSRHGYGETWQNRSSRSHTPIFMILESEKREKGPSPNCEGSKVYSHNLVD 234
 Db 201 QHRLHLSICEORQVGETALNTSSKSHQIILTVESRLR--EVSQCVKSF--VANLNFVD 258
 Qy 235 LAGSERAQAOTGAAGVRLKEGCNINRSFILGQVKKLSQGVGGFINYRDSKLTILQNS 294
 Db 259 LAGSERAQAOTGAAGVRLKEGCNINRSFILGQVKKLSQGVGGFINYRDSKLTILQNS 318
 Qy 295 LGGNPKTRIICTTPTPVF---DETALQFASAKYMKNTPYNEVSTDEALLKRYKEI 351
 Db 319 LGGNARTALICTMSPAQTHVQESQNTXLPFATCAKEVTNNAKVMVYSDQLVKHLQMEV 378
 Qy 352 MDLKKQLEEVSLLETRAQ--AMEKDOLAQLLEKDL--LOKVON---EKIENLTRL--- 400
 Db 379 ARLEAELRTPDRASSIIIIIMERDKIRQVKEKMEELKQORDNAQSKLEELQKMGDNQ 438
 Qy 401 -----VTSSSLTLOQELKAKRRRV 420
 Db 439 PGWNPEDSPQTRKCLTYSGSLQPSNKNKI 468

RESULT 11

US-10-311-642-2

; Sequence 2, Application US/10311642

; Publication No. US20040086878A1

; GENERAL INFORMATION:

; APPLICANT: Cytokinetics, Inc.

; APPLICANT: Beraud, Christophe

; APPLICANT: Freedman, Richard

; TITLE OF INVENTION: NOVEL MOTOR PROTEINS AND METHODS FOR THEIR

; FILE REFERENCE: 020552-001910US

; CURRENT APPLICATION NUMBER: US/10/311,642

; CURRENT FILING DATE: 2003-09-29

; PRIOR APPLICATION NUMBER: 09/597,602

; PRIOR FILING DATE: 2000-06-20

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 1029

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: HsKif17 amino acid sequence

; OTHER INFORMATION: Description of Artificial Sequence: HsKif17

US-10-311-642-2

Query Match 28.6%; Score 710.5; DB 16; Length 1029;
 Best Local Similarity 36.5%; Pred. No. 1.1e-51;
 Matches 190; Conservative 80; Mismatches 176; Indels 75; Gaps 16;

Qy 5 AVAVCVRVRLNSREESLGETAQVYWKTD-----NNVIYQVDSKSFNDFRVPHGNE 56
 Db 5 AVKVVVRCRPNQRELEL--RCQPWTVVDCARACQCIQNPQGADEPKQFTFDGAYVDH 62
 Qy 57 TTKNVVREIAPIDSALQGYNGTIFAYGQTASGKTYTMMGSED--HLGVIPIRAIHDI 113
 Db 63 VTEQIYNEIAYPLVEGTEGNGTIFAYGQTGSKGTFMQGLPDPSPORGHIIPRAFHFV 122
 Qy 114 QKIKKFPDRFLLRVSYMEIYNETITDLLCGTQKMKPLIREDVNRNRYVADLTTEVYVT 173
 Db 123 ESVQCAENTKFLVRASYLEIYNEVDRLDGADTKQK--LELKEHPEKGVYVVKLSMHTVHS 181
 Qy 174 SEMALKWITKEKSRHGYGETWQNRSSRSHTPIFMILESR---EKGPSNCEGSKVYSHL 230
 Db 182 VAQCEHINETWQNRSGYITLMKDSRSHTSIFTISIMSAVDERG-----KDLRACKL 236
 Qy 231 NLVDLAGSERAQAOTGAAGVRLKEGCNINRSFILGQVKKLSQGVGGFINYRDSKLTTRI 290
 Db 237 NLVDLAGSROSKTGATGERLKEATKINLSLSALGNVISALVDGRC--KHVYRDSKLTRL 295

Qy 291 LQNSLGNPKTRIICTTPTV--SFDETLTALQFASAKYMKNTPYNEVSTDEALLKRYR 348
 Db 296 LQDSLGGNTKTLMVACLSPADNNYDETLSTLRYANRAKNIRNKRINDEPKD--ALLREYQ 354
 Qy 349 KEIMDLKKQLEEVSLLETRAQAMEKDOLAQL-----BEKDLQKV----- 388
 Db 355 EEIKKKAIL-----TQOMSPSSLSALLSRQVPPDPVQVEBKLLPQVIOHMEAEK 406
 Qy 389 ---QNEKIENLTRLVTSSSLTLOQELKAKRRRVTCGLGKINKMKNKNYADQFNPTNI 445
 Db 407 QLRREYVEELARL---KADYKAEQSSRARLEEDIT-----AMRNSYDVRSLTEENL 456
 Qy 446 TTKTHKLSINLREIDESVCSDESVDPSNTLDTLSEIENPA 486
 Db 457 RKETEAV-----LQGVLYKAEVMSRA--EFASSAEYPPA 489

RESULT 12

US-10-116-712-664

; Sequence 664, Application US/10116712

; Publication No. US20030194764A1

; GENERAL INFORMATION:

; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: Switzer, Ann

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.568

; CURRENT APPLICATION NUMBER: US/10/116,712

; CURRENT FILING DATE: 2002-04-07

; NUMBER OF SEQ ID NOS: 670

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 664

; LENGTH: 1232

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-116-712-664

Query Match 28.3%; Score 702.5; DB 14; Length 1232;
 Best Local Similarity 35.1%; Pred. No. 6.9e-51;
 Matches 182; Conservative 86; Mismatches 191; Indels 59; Gaps 12;

Qy 6 VAVCVRVRLNSREESLG--ETAQVYWKTDNNVIYQVDSKSFNDFRVFHGNETTKNVYEE 64
 Db 10 VRVALRCRPLVPKEISEGQMCCLSFVDPGEQVVGTD--KSFYDFVDFDSTGEQEVNT 67
 Qy 65 IAAPIIDSALQGYNGTIFAYGQTASGKTYTMMG-----SEHLGVIPIRAIHDI 117
 Db 68 AVAPLKGVPKGYNATVAYGQTGSGKTYSMGGAYTAEQENEPVGVIPRVQLLFKRID 127
 Qy 118 KFPDRFLLRVSYMEIYNETITDLLCGTQKMKPLIREDVNRNRYVADLTTEVYVTSEMA 177
 Db 128 KKSDFEFTLVKSVYLEIYNEBILDLCPREKAQINIREDPKGIKIVGLTEKTVLVALDT 187
 Qy 178 LKMITKGEKSRHGYGETWQNRSSRSHTPIFMILESEKREKGPSPNCEGSKVYSHNLVDLAG 237
 Db 188 VSCLEQNGNRTVASTANNQSSRSHTAFTI----SLEQKKSKDNKSNFR--SKLHLVDLAG 243
 Qy 238 SERAAQGAAGVRLKEGCNINRSFILGQVKKLSQGVGGFINYRDSKLTILQNSLGG 297
 Db 244 SEROKTKAEGDRLKEGININRGLLCLGNVISALGDDKKGGFAYRDSKLTILQDLSGG 303
 Qy 298 NPKTRIICTTPTV--SFDETLTALQFASAKYMKNTPYNEVSTDEALLKRYKEIMDLK 355
 Db 304 NSHTLMIAVCSPADSNLEETLNTLRYADRARKINKKEIVN-----IDPQTAEHLNHLK 355
 Qy 356 KQLEEVSL-----ETRAQAMEKDOLAQLLEKDLQKQVQNEKIENL 396
 Db 356 QQVQLOQLVLLQAHGGTLPPOSITVPESENLOSLMEKQ--SLVEENKLSGLSEAAQT 413
 Qy 397 TRML---VTSSSLTLOQELKAKRRRVTCGLGKINK---MKNSNYADQFNPTNITTKT 449
 Db 414 AQMLERIITWTEQAENKKNNAKLEELRQHAACKLDLQKLVTLEDQELKENVEIICNLQOLI 473


```

; LENGTH: 672
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-763A-1664

Query Match      27.7%; Score 689; DB 16; Length 672;
Best Local Similarity 35.8%; Pred. No. 4e-50;
Matches 182; Conservative 81; Mismatches 151; Indels 94; Gaps 13;

Qy 31 KTDNNVIYQVDSKSNFDFVFGNETTKNVYEEIAPIIDSAIOGNGTIFAYGQTASG 90
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 20 KTDSS---NEPPKTFTFDFGPEKQLDVYNLTARPIIDSVLEGYNGTIFAYGQTG 75
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 91 KTYTMGSE---DHLGVIPIRAIHDIPOKIKKFP-DREFLLRVSYMEIYNETITDLLCGTQ 146
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 76 KTFMEGVRAIPERLGIPIPSPAHIFGHIKAEGRDFLRVSVLEIYNEEVRDLL-GKD 134
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 147 KMKPLIIRDNRRNVYVADLTEEVYVTSEMAIKWITKESRHYGETKMNQRSRSHTIF 206
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 135 QTORLEVKERPVGVIYIKDLSAYVNNADDMRIMTLGHKNRSVGATNMNEHSSRSHAIF 194
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 207 RMILESEKGEPSNCEGSKVSHLNVDLAGSRAAOTGAAGVRLKEGCNINRSLFLGQ 266
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 195 TITIECEKGIDGNMH--VRWGLHLVDLAGSERQAKTGATGQRLKEATKINLSLTGN 252
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 267 VIKKLSQGVGGFINYRDSKLTIRLQNSLGNPKTRIICITTPV--SFDETLTALQFAS 324
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 253 VISALVDGK-STHVPYRNSKLTRELLQDSLGNSTKMMCANIGPADYNYDETISTLRVANR 311
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 325 AKYMKNTPYVNEVSTDEALLKRYKEIMDLKKOLEE----- 360
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 312 AKNIKAKARINEDPKD-ALLFQFQKEIETELKKLEEGEISGSDISGSEEDDDDEGEVGE 370
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 361 -----VSLTRAQAMEKQOLAQLLE----EKD 383
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 371 DGEKKRRRIQIGKKVSPDKMIENQAKIDBERKALETKLDMEEEERNKAPAELEKREKD 430
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 384 LLOKVQNEKIENLTRLMLVTSSSLTIQOELKAKRRRVTTWICLGKINKMKNNSNYADQFNPT 443
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 431 LLKAAQOEHQ-----SLEKLSALEKKVI---VGGVDLLAKABEEQKLEES 473
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 444 NI-TTKTHKLSINLLREIDESVCSQSDV 470
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 474 NMELEERRKRAEQLRRELEEKQERLDI 501
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: July 29, 2004, 10:06:29
Job time : 25.324 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:28:21 ; Search time 7.27641 Seconds
(without alignments)
6437.961 Million cell updates/sec

Title: US-10-045-631b-88_COPY_2_488

Perfect score: 2483
Sequence: 1 AEEGAVAVCVVRPLNSREE.....SDVFSNTLTLSEIWNPAT 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2483	100.0	2563	1 S28261	centromere protein
2	1584	63.8	2954	2 T14156	kinesin-related pr
3	880	35.4	823	2 T52425	kinesin-like prote
4	844.5	34.0	888	2 D96619	protein T30E16.9 [
5	759	30.6	1459	2 T30196	kinesin motor prot
6	735.5	29.6	747	1 A57107	kinesin-related pr
7	728	29.3	701	1 B44259	kinesin-related pr
8	722	29.1	742	1 S58691	kinesin-related pr
9	716	28.8	786	2 A53939	kinesin homolog KH
10	710.5	28.6	1231	2 A54803	microtubule-associ
11	705	28.4	1121	2 T06065	hypothetical prote
12	704.5	28.4	699	1 S38982	kinesin-related pr
13	703	28.3	1066	1 A48669	kinesin-related pr
14	700	28.2	909	2 H86350	hypothetical prote
15	699	28.2	1130	2 T21134	hypothetical prote
16	694	28.0	581	2 F84599	probable kinesin h
17	692	27.9	1226	2 I51617	kinesin heavy chai
18	689.5	27.8	932	2 T49235	kinesin-like prote
19	687.5	27.7	968	2 T45746	hypothetical prote
20	676.5	27.2	1225	2 A56514	chromokinesin - ch
21	670	27.0	1263	2 T13465	hypothetical prote
22	664.5	26.8	1031	1 A34977	kinesin heavy chai
23	663.5	26.7	1254	2 T18277	kinesin heavy chai
24	660.5	26.6	975	1 A31497	kinesin heavy chai
25	656.5	26.4	834	2 T06055	hypothetical prote
26	656	26.4	672	2 S54351	kinesin osm-3 - Ca
27	654.5	26.4	963	1 A41919	kinesin heavy chai
28	652.5	26.3	1076	2 B84687	probable kinesin-l
29	652	26.3	793	2 JCS831	kinesin-related pr

30	650	26.2	1921	2 T13827	kinesin-73 - fruit
31	642.5	25.9	967	1 A35075	kinesin heavy chai
32	638.5	25.7	1032	2 I38510	neuronal kinesin h
33	635.5	25.6	881	2 I84737	kinesin heavy chai
34	635.5	25.6	968	2 T51933	kinesin motor prot
35	633	25.5	784	1 A55236	kinesin-related pr
36	632.5	25.5	1056	1 G02157	kinesin-like spind
37	631	25.4	1027	2 S37711	kinesin heavy chai
38	630.5	25.4	1067	2 S33417	kinesin-like prote
39	629	25.3	1060	1 A40264	kinesin-related pr
40	628	25.3	843	2 S44868	kinesin heavy chai
41	628	25.3	1388	2 T30335	KLP2 protein - Afr
42	624.5	25.2	857	2 E84600	probable kinesin h
43	624.5	25.2	1576	2 T29237	hypothetical prote
44	623.5	25.1	554	2 T50118	kinesin-related pr
45	622.5	25.1	1058	2 T47525	kinesin-related pr

ALIGNMENTS

RESULT 1
S28261
centromere protein E - human
N:Alternate names: centromere 312K protein; kinesin-related protein CENP-E
C:Species: Homo sapiens (man)
C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C:Accession: S28261
R:Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
Nature 359, 536-539, 1992
A:Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.
A:Reference number: S28261, MUID:93024922, PMID:1406971
A:Accession: S28261
A:Molecule type: mRNA
A:Residues: 1-2663 <YEN>
A:Cross-references: EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865
C:Genetics:
A:Gene: GDB:CENPE
A:Cross-references: GDB:361164; OMIM:117143
A:Map position: 4q24-q25
C:Superfamily: centromere protein E; kinesin motor domain homology
C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
F:7-335/Domain: kinesin motor domain homology <KWOT>
F:86-93/Region: nucleotide-binding motif A (P-loop)
F:486-2183/Domain: coiled coil #status predicted <COI>
F:92/Binding site: ATP (Lys) #status predicted

Query Match		100.0%;	Score 2483;	DB 1;	Length 2663;
Best Local Similarity		100.0%;	Pred. No. 1.2e-134;		
Matches 487;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	AEEGAVAVCVVRPLNSREESLG	TAQVYWKTDNNVYQVDGSKGFNDRVPHGNETTKN	60	
Db	2	AEEGAVAVCVVRPLNSREESLG	TAQVYWKTDNNVYQVDGSKGFNDRVPHGNETTKN	61	
Qy	61	VYEETAAPLIDSAIOGYNGTIF	FAYGQTASGKTYTWGSEDLGVIPRAIHDFQIKKPP	120	
Db	62	VYEEAAPLIDSAIOGYNGTIF	FAYGQTASGKTYTWGSEDLGVIPRAIHDFQIKKPP	121	
Qy	121	DREFLLRYSYMEIYANETITD	LLCGTKQMKPLIIRDVNRNVVADLTREVVYVTSSEALAKW	180	
Db	122	DREFLLRYSYMEIYANETITD	LLCGTKQMKPLIIRDVNRNVVADLTREVVYVTSSEALAKW	181	
Qy	181	ITKGSKSHYGETKMNQSSRSHT	IFRMLRESREKGEPSNCEGSKVKSHLNVLVLAGSR	240	
Db	182	ITKGSKSHYGETKMNQSSRSHT	IFRMLRESREKGEPSNCEGSKVKSHLNVLVLAGSR	241	
Qy	241	AAQTGAAGVRLKEGCNINRSL	FILGQVKKLSGQVGGFINRDRSKLTRILQNSLGGNPK	300	
Db	242	AAQTGAAGVRLKEGCNINRSL	FILGQVKKLSGQVGGFINRDRSKLTRILQNSLGGNPK	301	
Qy	301	TRIICTITPVSFDETLTALQF	ASTAKYMKNTPYNEVSTDEALLKRYBKEIMDLKKOLEE	360	

Db 302 TRIICITTPVSPDELTALQFASAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKQLEE 361
 QY 361 VSLTRQAQMEKQDLAGLLEKDLLOKVQNEKIENLFRMLVTSSTLTQQLKAKRRRV 420
 Db 362 VSLTRQAQMEKQDLAGLLEKDLLOKVQNEKIENLFRMLVTSSTLTQQLKAKRRRV 421
 QY 421 TWGLGKINKMNSYADQFNIPNTITTKHLSINLAREIDESVCSBDSFNTLDTLSE 480
 Db 422 TWGLGKINKMNSYADQFNIPNTITTKHLSINLAREIDESVCSBDSFNTLDTLSE 481
 QY 481 IEMNPAT 487
 Db 482 IEMNPAT 488

RESULT 2
 T14156
 kinesin-related protein - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
 C:Accession: T14156
 R:Wood, K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.
 Cell 91, 357-366, 1997
 A:Title: CNP-E is a plus end-directed kinetochore motor required for metaphase chromosome segregation
 A:Reference number: Z17893; PMID:98028574; PMID:9363944
 A:Accession: T14156
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2954 <WOO>
 A:Cross-references: EMBL:AF027728; NID:g2586070; PID:g2586071; PIDN:AAC60300.1
 C:Genetics:
 A:Gene: XCENP-E
 C:Superfamily: Centromere protein E; kinesin motor domain homology

Query Match 63.8%; Score 1584; DB 2; Length 2954;
 Best Local Similarity 64.0%; Pred. No. 7.4e-83;
 Matches 320; Conservative 68; Mismatches 96; Indels 16; Gaps 6;
 QY 1 AEEGAVCVVRPLNSREESLGETAQVYVKTNNVYQVDSKSFNFDRVPHGNETTKN 60
 Db 2 SEGDAVVCVVRPLNQREQ--GDOANLQWKAGNNTISQVDTGKSFNDRVFNHSTSQ 59
 QY 61 VVEEIAAPIIDSALQYNGTIFAYGQTASGKTYTMMGSEDLGLVPIRAIHDFQIKKFP 120
 Db 60 IYQELAVLIRLSALQYNGTIFAYGQTSSGKTYTMMGTPNSGLIPQAIQEVFKIIEIP 119
 QY 121 DREFLLRVSYMEIYNETITDLGCTQKMKPLIIRDVNRVYVADLTBEVVYTSMAKX 180
 Db 120 NREFLLRVSYMEIYNETITDLGCTQKMKPLIIRDVNRVYVADLTBEVVYTSMAKX 179
 QY 181 ITKGKSRHYGTQKNSRSRSHITFRMLTSEKGEPS---NCEGSKVVKSHLNLVDLAG 237
 Db 180 ITKGKSRHYGTQKNSRSRSHITFRMLTSEKGEPS---NCEGSKVVKSHLNLVDLAG 239
 QY 238 SERAAQTGAAGVRLKEGNCINRSLFILGQVIRKLSGQVGGFFINRYRDSKLTILQNSLGG 297
 Db 240 SERASQTGAEGVRLKEGNCINRSLFILGQVIRKLSGQVGGFFINRYRDSKLTILQNSLGG 299
 QY 298 NPKTRIICTITTPVSPDELTALQFASAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKQ 357
 Db 300 NAKTVIICITTPVSPDELTALQFASAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKQ 359
 QY 358 LE--SVSLTRQAQMEKQDLAGLLEKDLLOKVQNEKIENLFRMLVTSSTLTQQLKAK 415
 Db 360 LENLSSSTKQAQMAKEHTQLLAIKQLHREDRIWHLINIVASSQES--QDQRYK 418
 QY 416 RKRRTVWCLGKINKMNSYADQFN----IPNTITTKHLSINLAREIDESVCSBDSVF 471
 Db 419 RKRRTVWCLGKINKMNSYADQFN----IPNTITTKHLSINLAREIDESVCSBDSVF 478
 QY 472 SNTLDTLS----EIEWNPAT 487
 Db 479 DDALSMDSNGIDAENNLAS 498

RESULT 3

T52425
 kinesin-like protein [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000
 C:Accession: T52425
 R:Kato, A.; Suzuki, M.; Kuwahara, A.; Ooe, H.; Higano-Inaba, K.; Kameda, Y.
 Gene 239, 309-316, 1999
 A:Title: Isolation and analysis of cDNA within a 300 kb Arabidopsis thaliana genomic re
 A:Reference number: Z25171
 A:Accession: T52425
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-823 <KAT>
 A:Cross-references: EMBL:AB028468; PIDN:BAA88112.1
 C:Genetics:
 A:Gene: ZCF125

Query Match 35.4%; Score 880; DB 2; Length 823;
 Best Local Similarity 43.1%; Pred. No. 4.4e-43;
 Matches 220; Conservative 70; Mismatches 166; Indels 54; Gaps 12;
 QY 6 VAVCVVRPLNSREESLGETAQVYVKTNNVYQVDSKSFN-----FDRVPHGNE 56
 Db 4 ICVAVVRP-----PAPENGASLWKVEDN--RISLHKSJLDTPIITASHAPDHVDFDESS 54
 QY 57 TTKNVVEIAAPIIDSALQYNGTIFAYGQTASGKTYTMMGSEDLGLVPIRAIHDFOKI 116
 Db 55 TNASVYELLTKDIIHAAVEGFGNFAFGQTSSGKTFMTGSETDPGIIIRSVRDVPERI 114
 QY 117 KFPDREFLLRVSYMEIYNETITDLGCTQKMKPLIIRDVNRVYVADLTBEVVYTSSEM 176
 Db 115 HWSIDREFLLRVSYMEIYNEBNDLL--AVENQRLQIHEHLERGTVFVAGLKEEIVSDAEQ 172
 QY 177 ALKWITGKSRHYGTQKNSRSRSHITFRMLTSEKGEPSNCEGSKVVKSHLNLVDLA 236
 Db 173 ILKLIDSGEVNRHFGETNMNVHSSRSHITFRMVIESR--GKDNSSSDAIRVSVLNLVDLA 230
 QY 237 GSERAAQTGAAGVRLKEGNCINRSLFILGQVIRKLSGQVGGFFINRYRDSKLTILQNSL 295
 Db 231 GSERIAKTGAGVRLQEGKYINKSLMILGNVINKLSDSTKLRAHIPYRDSKLTILQPAL 290
 QY 296 GGNPKTRIICTITTPVS--FDELTALQFASAKYMKNTPYVNEVSTDEALLKRYRKEIMD 353
 Db 291 GGNAKTCIICITIAPEHHIESKGLTQFASAKRITNCAQVNEILTDALLKROKLEIEE 350
 QY 354 LKKQLEEVSLTRAQ-----AMXKDLQALLLEKDLLOKVQNE-----KIE 394
 Db 351 LRMLQGGSHAQVLEQELNLNLSNQLKYELECEERLTKLEEEKRKEQENCIKEQOMKIE 410
 QY 395 NLTRMLVTSSTLTQQLKAKRRRVTWCLGKINKMNSYADQFNIPNTITTKHLSI 454
 Db 411 NLNN--FVNSDF-----KRNQSEDFIISKTPDGLCNVNDTSDVPTGTCFKSASRF 461
 QY 455 NLLREIDESVCSBDSVFNTL--DTLSEIEW 483
 Db 462 VVARSNNSYGLSDFSPVHSLGDAVEDTW 491

RESULT 4

D96619
 protein T30B16.9 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: D96619
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.
 ansen, N.F.; Hughes, B.; Huijzer, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96619
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-888 <STO>
A:Cross-references: GB:AE005173; NID:g8778739; PIDN:AAF79747.1; GSPDB:GN00141
C:Genetics.
A:Gene: T30E16.9
A:Map position: 1.

Query Match.	34.0%;	Score 844.5;	DB 2;	Length 888;
Best Local Similarity	40.0%;	Pred. No. 5.4e-41;		
Matches	220;	Conservative	169;	Indels 91; Gaps 13;
QY	6	VAVCVVRPLNSREESIGETAQVYWKTDNNVI	-----Y 38	
Db	4	ICVAVVRP-----PAPENGASLWKVEDNRISLHKS	LDTPITPASHAFVSGISISTDLI 57	
QY	39	QVDGSKSFNF-----DRVPHGNETTKNVYEEIAAP	FLIDSAIOGYNGTIFAYGQTA 88	
Db	58	EIVSLFLFSGVVYFFLLPADHVFDESSWASVYELLTK	DIIRHRAVEGFNGTAPAYGQTS 117	
QY	89	SGKTYTMGSEDLHGVIPRAIHDIFOKIKFPDREF	LLRVSYMEIYNETITDLLCGTQM 148	
Db	118	SGKITFTMGSETDPGIIRRSVRDVFRIHMSDREF	LIRVSYMEIYNEEINDLL--AVEN 175	
QY	149	KPLIIREDVNRNVVYADLTVEVVYVTSSEMALKW	ITKGEKSHYGETWKNQBSSSHIFRM 208	
Db	176	ORLOIHEHLERGVPVAGLKEEIVSDAQILKLD	SGEVNHFGETNNVHSSRSHIFRM 235	
QY	209	-----ILSREKGEPSNCBGSVKVSHNLVLD	LAGSBRAAQCTGAAGVRLKEGCN 256	
Db	236	VFRFSYERDLLLVIESR--GKONSSDAIRSV	LVNLVDLAGSERIAKTGAGVRLQEGKY 293	
QY	257	INRSFILLQGVIKKLSDG-QVGGFNYR	OSKTRILQNSLGNPKTRIICTIPVS--FD 313	
Db	294	INKSIMILGNVINKLSDSTKLRAHIPYRDSK	TRILQPALGNAKTCIICTIAPEEHIE 353	
QY	314	ETLTALQFASTAKYMKNTPYVNEVSTDEALL	KYRKEIMDLKKQLEVSLEITPAQ----- 368	
Db	354	ESKGTGLQFASRAKRIITCAQVNEILTDAALL	KQKLEIEELRWKLOGSHAEVLEQIILNL 413	
QY	369	-----AMEKDQLAQLBEKOLLQKVQNE-----	KIENLTRMLVTSSSLTLOQLKXA 414	
Db	414	SNQMLKYELCEERLKTQLEEEERKQKEQNC	IEKQOMKIENLNN-FVITNSDF----- 464	
QY	415	KRKRRVTVCLGKINKMKNYSYADQENIP	NIITKTHKLSINLLREIDESCYESDVFSWT 474	
Db	465	KENQSEDFIISKTPDGLCNVNDTSDVP	GTPCFKSASRFVVARNNYSGLDFSPMVHS 524	
QY	475	L-DTLSETEW 483		
Db	525	LGDVADEDTW 534		

RESULT 5
T30196
kinesin motor protein 1 - smut fungus (*Ustilago maydis*)
C.Species: *Ustilago maydis* (corn smut)
C.Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C.Accession: T30196
R.Lehmler, C.; Steinberg, G.; Snetselaar, K.M.; Schliwa, M.; Kahmann, R.; Bolker, M.
EMBO J. 16, 3464-3473, 1997
A.Title: Identification of a motor protein required for filamentous growth in *Ustilago*
A.Reference number: Z20770; MUID:97361828; PMID:9218789
A.Accession: T30196
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-1459 <LEH>

A;Cross-references: EMBL:U92844; NID:g2062749; PID:g2062750; PIDN:AAB63336.1
C;Genetics:
A;Gene: kin1
C;Function:
A;Description: required for filamentous growth in *Ustilago maydis*

Query Match 30.6%; Score 759; DB 2; Length 1459;
Best Local Similarity 38.2%; Pred. No. 8.7e-36;
Matches 191; Conservative 71; Mismatches 128; Indels 110; Gaps 14;

Query Match	30.6%	Score 759;	DB 2;	Length 1459;
Best Local Similarity	38.2%;	Prod. No. 8.7e-36;		
Matches 191;	Conservative 71;	Mismatches 128;	Indels 110;	Gaps 14;

QY	6	VAVCVVRPLNSREESIGETAYWKTDN--NVYQVD-----	41
DB		:	
QY	237	VVVCVRMP--SRASSDSSEASY--WNCDSSEKRIFFTEHHPALAKRTTSSERAGAGASIA	293
DB		:	
QY	42	-----GSKSFNDFRVFHNETHKNVYEEETAAPIIDSAIOGYNGTIFAYGQ	86
DB		:	
QY	294	AAPSSHDLDHEDPTSTSYHFQDKLTGTGAQTTDDMVHSHIAPVRAAVSGYNGTIVFAYGQ	353
DB		:	
QY	87	TASGKTYTWMGSEDHGLGVIPTRAIHDFQIKKPPDFRELLRVSYMEIYVNETITDLCGQTQ	146
DB		:	
QY	354	TGSGKTHTMGSGDAEPGVIPRAVEQIFQMKDEPREFLLRVSYLEIYNETLKLLA---	410
DB		:	
QY	147	KMKPL-----	
DB		:	
QY	411	PIPLPTGSSSLQTTDRPASPICKGSSHAAGSQSCTLIREDOKSSRVITGLREEV	470
DB		:	
QY	172	YTESMALKWITKEGSRHYGETKMNORSRSHTIIFRMILESEKGEPSCEGSVKVXSHLN	231
DB		:	
QY	471	TDANTVLCIJQRQDERHVGATDNWERSRSRCHVQLTIESRSPAPSASKE--VRISQLN	528
DB		:	
QY	232	LVDLAGSERRAOTGAAGVRLKEGCNINRSLFILGQVIKLSDGQVG--FINYRDSKLTR	289
DB		:	
QY	529	LIDLAGSERAA--SQAERKEGAFINKSLLTLGTIVIGKLTPEVNGDAHIPIYRDSKLTR	585
DB		:	
QY	290	IIQNSLGNPKTRIICITTPVS--FDELTALQFASTAKYMKNTPYVNEVSTDEALLKRY	347
DB		:	
QY	586	IIQTSLSGNARNAVICTISPDTEHANETLSTLFGKRCKLVVYTTAKGTGTMDDKALLOKY	645
DB		:	
QY	348	RKEIMDLKKQLEE-----VSLETRAQAAMEK-DQLAOLLE-----EKDLLQKVQ	389
DB		:	
QY	646	RKELDALRAKLEANGSPNPNEVMTIVSABASKESQKQLDQUNQOKEAARQREVEDMOKKR	705
DB		:	
QY	390	NE---KIENITRMVLVTSSSL	406
DB		:	
QY	706	SHLKAQIEHLTELILTSQSV	725
DB		:	

RESULT 6

A57107
kinesin-related protein KIF3B - mouse
C:Species: Mus musculus (house mouse)
C:Date: 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change 19-Jan-2001
C:Accession: A57107
R:Yamazaki, H.; Nakata, T.; Okada, Y.; Hirokawa, N.
J. Cell Biol. 130, 1387-1399, 1995
A:Title: KIF3A/B, a heterodimeric kinesin superfamily protein that works as a microtubul
A:Reference number: A57107; MUID:96032268; PMID:7559760
A:Accession: A57107
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-747 <YAM>
A:A:Cross-references: GB:D26077; NID:g1060922; PIDN:BAA05070.1; PID:g1060923
A:Experimental source: brain
C:Complex: heterodimer with KIF3A (PIR:B44259); the KIF3A/3B heterodimer associates with
A:Description: KIF3 complex is a motor protein that provides anterograde fast axonal tra
C:Superfamily: kinesin-related protein KIF3; kinesin motor domain homology
C:Keywords: ATP; coiled coil; heterodimer; heterotrimer; microtubule binding; nucleotide
F11-363/Domain: head globular #status predicted <HGL>
F10-346/Domain: kinesin motor domain homology <RMOT>
F10-103/Region: nucleotide-binding motif A (p-loop)
F1364-592/Domain: helical rod #status predicted <ROD>
F1594-747/Domain: tail globular #status predicted <TGL>

F;102/Binding site: ATP (lys) #status predicted

Query Match 29.6%; Score 735.5; DB 1; Length 747;
 Best Local Similarity 37.9%; Pred. No. 8.1e-35;
 Matches 193; Conservative 85; Mismatches 180; Indels 51; Gaps 13;

QY 5 AVAVCVVRPLNRSRESLG-----ETAQVYVWKTDNNVVIQVQDGSKFNFDRVPHGN 55
 DB 9 SVRVVRCRPMNGKGEKAASVYKVDVVDVVKLGQVSVKPKGTSHM--PKTFTDAVVDWN 66
 QY 56 ETTKNVVEELAAPIIDSAIOGYNGTIFAYGQTASGKTYTMM--MGSEDHILGVIPRAIHDI 112
 DB 67 AKQELVDETRFLVDSVLQGFNGTIFAYGQTGKTYTMEGVGRDPEKRGVIPSFDHI 126
 QY 113 FQIKKPPDRFLRVSYMEIYNTITDLCGTQKMKPLIREDVNRNVVADLTERVY 172
 DB 127 FTHISRQOQYLVRASYLEIYQBEIRDLLSKDQ--TKRELKERPDPTGVVYVKDLSSFVTK 185
 QY 173 TSEMALKWITKESRHYGETKMNORSRSHITFRMILESEKGEPCNCEGSVKVSHLN 232
 DB 186 SVKEIEHVMNVGNQNRSGVATNMHESSRSHAFVITIECSEVG--LDGENHIRVGKLN 243
 QY 233 VDLAGSRAAQTGAAGVRLKEGNCINRSLFILGVQVVKLSQGVGFNINRDSKLTLLQ 292
 DB 244 VDLAGSERQAKTGAGGERLKEATKINLSLGNVISALVDGK--STHIPYRDSKLTLLQ 302
 QY 293 NSLGNPKRICTITPVSF--DETILTAQFASAKYMKNTPYVNEVSTDEALLKRYRKE 350
 DB 303 DSLGNGAKTVMVANVPASVYNEETLTILRYANRAKNIKPRVNEDEPKD--ALLREFQEE 361
 QY 351 IMDLKKOLEEVSLETRAQ-----AMERKQDLAQLLEE---KDLQKQVQNEKI 393
 DB 362 IARLKAQLEKRSIGRRKREKREGGGGGGEEEGEEDDDKDDYWRQEQEKL 421
 QY 394 ENLTRLMTVSSSLITLQELK-----AKRKRVTWCLG-KINKMKNVSNYADQNI 441
 DB 422 ELEKRAIVEDHSLVAEKMRLLKEKEMEDLRREKDAEMLGAKIKAMESKLLVGGKNI 481
 QY 442 PNITTKYHKSINLLREIDESVCSQSDV 470
 DB 482 -VDHTNEQKILEQKREIAEQRREREI 509

RESULT 7

B44259
 Kinesin-related protein KIF3A - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
 C:Accession: B44259; S27872
 R:Alzawa, H.; Sekine, Y.; Takemura, R.; Zhang, Z.; Nangaku, M.; Hirokawa, N.
 J. Cell Biol. 119, 1287-1296, 1992
 A:Title: Kinesin family in murine central nervous system.
 A:Reference number: A44259; MUID:93077686; PMID:1447303
 A:Accession: B44259
 A:Molecule type: mRNA
 A:Residues: 1-701 <AIZ>
 A:Cross-references: EMBL:D12645; NID:g220469; PIDN:BAA02166.1; PID:g220470
 A:Experimental source: brain
 A>Note: sequence extracted from NCBI backbone (NCBIP:118911)
 C:Complex: heterodimer with KIF3B (PIR:A57107); the KIF3A/3B heterodimer associates with C;Function:
 A:Description: KIF3 complex is a motor protein that provides anterograde fast axonal tra
 C:Superfamily: kinesin-related protein KIF3; kinesin motor domain homology
 C:Keywords: ATP; coiled coil; heterodimer; heterotrimer; microtubule binding; nucleotide
 F;1-368/Domain: head globular #status predicted <HGL>
 F;15-351/Domain: kinesin motor domain homology <KMT>
 F;100-107/Region: nucleotide-binding motif A (P-loop)
 F;369-599/Domain: helical rod #status predicted <ROD>
 F;600-701/Domain: tail globular #status predicted <TGL>
 F;106/Binding site: ATP (lys) #status predicted

Query Match 29.3%; Score 728; DB 1; Length 701;
 Best Local Similarity 41.7%; Pred. No. 2e-34;

Matches 184; Conservative 72; Mismatches 149; Indels 36; Gaps 11;

QY 6 VAVCVVRPLNRSRESL--GETAQVYVWKTDNNVVIQVQDGS-----KSNFDRVPHGNETT 59
 DB 15 KVWVRCRPLNREKSMCYRQAVSDVMRGTTIVHKTDSSNEPPKFTPTDVFQPSKQL 74
 QY 60 NVVEELAAPIIDSAIOGYNGTIFAYGQTASGKTYTMMSE---DHLGVIPIRAIHDFQKI 116
 DB 75 DVYNLTARPIIDSVLEGNGTIFAYGQTGKTYTMEGVRAVPCLRGVIPNSFAHIFGHI 134
 QY 117 KKEP--DREPLRVSYMEIYNTITDLCGTQKMKPLIREDVNRNVVADLTERVYVTS 175
 DB 135 AKAEGRTRFLRVSYLEIYNEEVRDLL-GKDQ*QRLVEKRPDVGVIKOLSAVVVNAD 193
 QY 176 MALKWITKESRHYGETKMNORSRSHITFRMILESEKGEPCNCEGSVKVSHLNLDL 235
 DB 194 DMDRIMTLGHKRSVGNATNMHESSRSHAFITIECEKGVQDNMH--VAMGKLHLVDL 251
 QY 236 AGSERAQCTGAAGVRLKEGNCINRSLFILGVQVVKLSQGVGFNINRDSKLTLLQNSL 295
 DB 252 AGSERQAKTGATGQRLKEATKINLSLGNVISALVDGK--STHVPYRNSKLTLLQDSL 310
 QY 296 GGNPKRICTITPV--SFDLTALQFASAKYMKNTPYVNEVSTDEALLKRYRKEIMD 353
 DB 311 GNSKTMWCANIGADYNYDETISTLYANRAKNIKPRVNEDEPKD--ALLRQFQKEIE 369
 QY 354 LKQJL---EVSLETRAQAMEKQDLAQLLEEKDLQKQNE-----KIE 394
 DB 370 LKKLEEGEEVSGSDISGSEEDDEGELGEGEKRRKRRDQAGKKKVSFDMKVMQAKID 429
 QY 395 NLTRMLVTSSTLITLQELKAK 415
 DB 430 BERKALETKLDMEERKAR 450

RESULT 8

S58691
 Kinesin-related protein KRP95 - sea urchin (Strongylocentrotus droebachiensis)
 A:Alternate names: kinesin-2 chain B; KRP (85/95) 95K chain
 C:Species: Strongylocentrotus droebachiensis
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
 C:Accession: S58691
 R:Rashid, D.J.; Wedaman, K.P.; Scholey, J.M.
 J. Mol. Biol. 252, 157-162, 1995
 A:Title: Heterodimerization of the two motor subunits of the heterotrimeric kinesin, KRP95
 A:Reference number: S58691; MUID:95404610; PMID:7674298
 A:Accession: S58691
 A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
 A:Molecule type: mRNA
 A:Residues: 1-742 <RAS>
 C:Complex: heterotrimer of a 115K chain and two kinesin-related chains of 85K (PIR:S38959)
 C:Superfamily: kinesin-related protein KIF3; kinesin motor domain homology
 C:Keywords: ATP; heterotrimer; microtubule binding; nucleotide binding; P-loop
 F;9-345/Domain: kinesin motor domain homology <KMT>
 F;95-102/Region: nucleotide-binding motif A (P-loop)
 F;101/Binding site: ATP (lys) #status predicted

Query Match 29.1%; Score 722; DB 1; Length 742;
 Best Local Similarity 37.6%; Pred. No. 4.8e-34;
 Matches 198; Conservative 74; Mismatches 191; Indels 64; Gaps 13;

QY 6 VAVCVVRPLNRSRESLGETAQVYVWKTDNNVVIQV-----DGSKSNFDRVPHGNETT 58
 DB 9 KVWVRCRPMNSKEISQGHKRIIVEMDNKRLVEVTNPKGPPGPNKSFDTVDVWNSKQ 68
 QY 59 KNVVEELAAPIIDSAIOGYNGTIFAYGQTASGKTYTMMG---SEDHILGVIPRAIHDFQK 115
 DB 69 IDLYDETRFLRVSYMEIYNTITDLCGTQKMKPLIREDVNRNVVADLTERVYVTS 128
 QY 116 IKKPPDRFLRVSYMEIYNTITDLCGTQKMKPLIREDVNRNVVADLTERVYVTS 175
 DB 129 IARTNQOFLVRASYLEIYQBEIRDLLAKDQK--KRLDLKERPDPTGVVYVKDLSSFVTSVK 187

355 KQLEEVSL-----ETRAQAMEKDQLAQLEEKDLQKQVONEKIN 395
 356 KQVQOQLOILLQAHGCTLPDGINVSENLQSLMEKQ--SLVEENEKLSRGLSEAAQ 413
 396 LTRML---VTSSLTLOELKAKRKRVTWCLGKINK-----MKNSYADQFNPTNTTK 448
 414 TAQMLERIILTEQANERKNNAKLEELRRAHAAKVDLQKLVTLEDOELKENIEICNLQQV 473
 449 THKLS-----INLLRIBESVCS-----ESDVFS 472
 474 IAQLSDEAAACMTATIDTAGEADTVQSSPDTGRSSDVFS 513

RESULT 11
 T06065
 Hypothetical protein F19H22.150 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 31-Mar-2000
 C:Accession: T06065
 R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X
 submitted to the Protein Sequence Database, March 1999
 A:Reference number: Z15184
 A:Accession: T06065
 A:Molecule type: DNA
 A:Residues: 1-1121 <BEV>
 A:Cross-references: EMBL:AL035679; GSPDB:GN00062; ATSP:F19H22.150
 A:Experimental source: cultivar Columbia; BAC clone F19H22
 C:Genetics:
 A:Gene: ATSP:F19H22.150
 A:Map position: 4
 A:Introns: 139/2; 170/1; 200/1; 256/3; 284/3; 302/3; 322/3; 349/1; 371/3; 390/3; 425/3;
 C:Superfamily: kinesin heavy chain; kinesin motor domain homology
 F:99-494/Domain: kinesin motor domain homology <KMT>

Query Match 28.4%; Score 705; DB 2; Length 1121;
 Best Local Similarity 34.3%; Pred. No. 7.8e-33;
 Matches 196; Conservative 96; Mismatches 148; Indels 132; Gaps 18;

QY 1 AEEGAVCVVRPLN-----SREES-----LGETA---QVYWKTDN 34
 DB 94 SERDSISVTRFRLPYARSPLAMDQIYCRDYSFHDVAIGVNSLLGQDSFGIPLAIDVF 153
 QY 35 NVI--YQVDSKSNFVRPHGNETTKNVEIAAIPDSAIQGYNGTIFAYGOTASGKT 92
 DB 154 DTLVREYNPLTAFAFKVQPGQATIDVDVAPVPVKAAMEGVNGVFAVGYTSSGKT 213
 QY 93 YTM-----MGSEDLHLGVIPRAIHDF 113
 DB 214 HTMHRVRLKKNPYTPFSLVFLQNTCAMNWDLPFNVLPLWLLGQDSFGIPLAIDVF 273
 QY 114 QKIKF-----PDREFLRYSYMEIYNETITDLCGQTKMKPLIREDVNRNVVADLT 167
 DB 274 SIQDVSLNGTGPREFLRYSYMEIYNEVINLDDPTG--QNLRVED--SQGYVEGK 330
 QY 168 EEVYVTSMAKWTITGKSRHYGETKQNRSSRSHITIFRMILSRKSGPSNCEGSVKV 227
 DB 331 EEVVLSPGHALSFAAGEEHRHVGSNFNLLSSRSHITFTLWVSSATGDEVD---GVIF 387
 QY 228 SHLNIIVLAGSRAAQTGAAGVRLKEGNCINRSIFILQVVKKLSGQGVGFNYRDSKL 287
 DB 388 SQLNLIDLAGSE--SSKTTTGLRKEGSIYKNSLLTLGTIVKLSGK--ATHIPIYRDSL 445
 QY 288 TRILONSIGGNPKTRITCTIPV--SPDETTLAQFASTAKYMKNTPYVNEVSTDEALLK 345
 DB 446 TRLQSSLSHGHSVLCITITPASSSEETNLIKFSRKSIEIYASRNIIDEKSLIK 505
 QY 346 RYRKEIMDLKKOLEEV-----SLETRAQAMEKDQ--LAQLLEEKDLQKQVONEK 392
 DB 506 KYQREISTKLDELQRLRGMLVGVSHLEMSLKQLEEGQVKMQSRLEEEBEEBAAALMSR 565
 QY 393 IENLTRMLVTSLSLTLOELKAKRKRVTWCLGKINKMNSYADQFNPTNTTKTKL 452

566 IQKTLKILVST-----KNSIFGYSGDIP-----THQR 593
 QY 453 SINLLREID-ESVCSSESDVF--SNTLDTLSE 480
 DB 594 SLSAGKDRFDSLLLESNDNLGSPSSITALLSE 625

RESULT 12
 S38982
 kinesin-related protein KRP85 - sea urchin (Strongylocentrotus purpuratus)
 N:Alternate names: kinesin-2 chain A; KRP (85/95) 85K chain
 C:Species: Strongylocentrotus purpuratus (purple urchin)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
 C:Accession: S38982; S72551
 R:Cole, D.G.; Chinn, S.W.; Wedaman, K.P.; Hall, K.; Vuong, T.; Scholey, J.M.
 Nature 366, 268-270, 1993
 A:Title: Novel heterotrimeric kinesin-related protein purified from sea urchin eggs.
 A:Reference number: S38982; MUID:94050179; PMID:8222586
 A:Molecule type: mRNA
 A:Residues: 1-699 <COL1>
 A:Cross-references: EMBL:L16993; NID:g295245; PIDN:AAA16098.1; PID:g295246
 A:Accession: S72551
 A:Molecule type: protein
 A:Residues: 2-5, 'X', 7-11, 59-64, 125-132, 222-226, 'X', 228-230 <COL2>
 C:Complex: heterotrimer of a 115K chain and two kinesin-related chains of 95K
 C:Superfamily: kinesin-related protein KIF3; kinesin motor domain homology
 C:Keywords: Atp; heterotrimer; microtubule binding; nucleotide binding; P-loop
 F:11-348/Domain: kinesin motor domain homology <KMT>
 F:97-104/Region: nucleotide-binding motif A (P-loop)
 F:103/Binding site: ATP (Lys) #status predicted

Query Match 28.4%; Score 704.5; DB 1; Length 699;
 Best Local Similarity 36.2%; Pred. No. 4.5e-33;
 Matches 200; Conservative 76; Mismatches 152; Indels 125; Gaps 19;

QY 6 VAVCVVRPLNRSRESLG-----ETAQYWKTDNNVIYQVDSKSNFDRPHG 54
 DB 11 VRVVRCPRLNSKETGQFKSVVXMDMRGTQV---TNPNA-PSGPPKSFDTFVAP 66
 QY 55 NEITKNVYEIAAIPDSAIQGYNGTIFAYGOTASGKTYTMWG--SEDHL-GVIPRAIHD 111
 DB 67 GAKQTDVYNQATARPIDVDAIEGNGTIFAYGOTGCTGKFTMEGVRSPQLGIPLNSFAH 126
 QY 112 IFQKIKFPDR-BFLRVSYMEIYNETITDLCGQTKMKPLIREDVNRNVVADLTVEV 170
 DB 127 IFGHIAKEQENRFLVRVSYLEIYNEEVKDLL-GKQDQHRLEVKERPDVGVYVKDLGSAFV 185
 QY 171 VYTSMAKWTITGKSRHYGETKQNRSSRSHITIFRMILSRKSGPSNCEGSVKVSHL 230
 DB 186 VNNADDMDRITMTLGNKRSVGATNMNNESSSRSHAFTITLERSDMG--LDKEQHRVRGKL 243
 QY 231 NLVDLAGSRAAQTGAAGVRLKEGNCINRSIFILQVVKKLSGQGVGFNYRDSKLTRI 290
 DB 244 HMVDLASEGRTKTCATGQRLKEATKLNLSLTGNSLVLDCK-SHIPIYRNSKLTRL 302
 QY 291 LQNSLGNPKTRITCTIPV--SPDETTLAQFASTAKYMKNTPYVNEVSTDEALLKRYR 348
 DB 303 LQDSLGNNAKTVMCANIGPAEYNYDETISTLRVANRAKNIKAKINAKINEDPKD-ALLREFQ 361
 QY 349 KEIMDLKKOLEE-----VSLETRAQAM 370
 DB 362 KEIEELKKQISESGEGLDDEESGESBEGEAGEGVKKRKKGNPKRKLSPKPEIMA-AM 420
 QY 371 EK--DQLAQLEE-KDLL-----QKVNQEKIENLTRMLVT- 402
 DB 421 QKKTDEBKALEEKQDVEEDRNTVHRELQRESELHKAQDDQKILNELKNAIQKLIIVG 480
 QY 403 -----SSSLTLQOELKAKRKRVTWCLGKINKMNSYADQFNPTNTTK 446
 DB 481 GVDLLAKSEQELLESQSALEMKERMAKQESMR-----KMMEEERQERMDIEEKYS 531
 QY 447 T-----KTHKL 452

CJ:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T21134

R:Matthews, P.

submitted to the EMBL Data Library, December 1995

A:Reference number: Z19381

A:Accession: T21134

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1130 <WIL>

A:Cross-references: EMBL:Z68161; PIDN:CAA92295.1; GSPDB:GN00022; CESP:F20C5.2

A:Experimental source: clone F20C5

C:Genetics:

A:Gene: CESP:F20C5.2

A:Map position: 4

A:Introns: 15/3; 34/2; 69/2; 152/3; 183/3; 207/2; 314/3; 513/3; 546/2; 594/1; 632/3; 656/2

Query Match 28.2%; Score 699; DB 2; Length 1130;

Best Local Similarity 34.8%; Pred. No. 1.8e-32;

Matches 194; Conservative 87; Mismatches 174; Indels 102; Gaps 16;

QY 1 AEEGAVAVCVRPPLNSREBSLGETAQVYKWTNNVI-----YQVDGSKSFNFDVPHG 54

DB 9 SKQETVKVIVRCPLSQSEIANNYSKIVHWRPQGQIELKNPKQEOPSKDFDIDAYDE 68

QY 55 NETTKNVEEIAPIIDSAIQGVNGTTFAYGQTAGSKTYTMGSE----DHLGVIPRAIHD 111

DB 69 NSTQSDIYEETFRDLVDVSLNGYNATTFAYGQTGTGKTHMEGKSDPEQRGVYKCIDH 128

QY 112 IFOKIKFPDPREFLLRVSYMEIYNETITDLLCGTQKMKPLIREDVNVYVADLTVEV 171

DB 129 IFEHMAASHNQEYLVASYLEIYQEEELRDL--EAESNKKLEIKERPDPGVVVKOLTSKIT 187

QY 172 YTEMALKWITKGEKSHYGETKNORSSRSHTTFRMLESREKGEPSNCEGSVKVSHLN 231

DB 188 RTVGEIHEVMIRGNCHRSVGRKTNWNEHSSRSHAFITVECSRIGEDG--BSHITVGRLN 245

QY 232 LVDLAGSERAQAQGAAGVRLKEGCNINRSLFILGOVIKKLSDGQGVGGHINTRDSKLTIRL 291

DB 246 LVDLAGSERQSKTGATGERFKEATKINLSALGNVISALVDK--SAHIPYRDSKLTIRLL 304

QY 292 QNSLGGNPKTRILCTIIPVSPDETLTALQFASKYMENTPYVNEVSTDEALLKRYRKEI 351

DB 305 QDSJGGNSKT-----ETLGTURYANRKNIKNQPKINEDPKD--ALLREFQEEI 351

QY 352 MDLAKQQL-----EEVSLTEFAQM-----EKDQIAQLLEEK-D 383

DB 352 EMLREQLKQRTRSRDQATQSFYDAERAKLEDDIEAIQKDDSLIKHEDRLTIREIQEKHD 411

QY 384 LQKQVQ-----NEKIENLTMILVTSS-----SLTLQOELKAKRERVTCWLGKINKM 430

DB 412 LLEKERIEQARVAERIAIQRLIVGSBEDGRLEBSRTKEQHAQLEKKR-----ELAEQ 465

QY 431 K-----NSNDAOFNPITNTTTTKHLSINLLR-----BIDBSV 464

DB 466 KRREREMVEALERQESDTVDLQKTFSD---LRTEVEAKTKKLKMWLIKLRQARNEIRDSV 522

QY 465 CSESDVFNLTDLSEI 481

DB 523 GAYSDERQDLDTIAEV 539

Search completed: July 29, 2004, 09:41:06
Job time : 8.27641 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:26:36 ; Search time 4.67769 Seconds
(without alignments)
5421.082 Million cell updates/sec

Title: US-10-045-631b-88_COPY_2_488

Perfect score: 2483

Sequence: 1 AEEGAVAVCVRRVPLNSREE.....SDVFSNTLDTLSEIEWNPAT 487

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2483	100.0	2663	1	CENE_HUMAN	Q02224 homo sapien
2	735.5	29.7	747	1	KF3B_HUMAN	O15066 homo sapien
3	735.5	29.6	747	1	KF3B_MOUSE	Q61771 mus musculus
4	728	29.3	701	1	KF3A_MOUSE	P28741 mus musculus
5	723	29.1	702	1	KF3A_HUMAN	Q9V496 homo sapien
6	722	29.1	742	1	KI21_STRPU	P48871 strongyloce
7	716	28.8	786	1	FL10_CHLRE	P48869 chlamydomon
8	713.5	28.7	1232	1	KF4A_HUMAN	O95239 homo sapien
9	710.5	28.6	1029	1	KF17_HUMAN	Q9P2e2 homo sapien
10	710.5	28.6	1231	1	KF4A_MOUSE	P33174 mus musculus
11	704.5	28.4	699	1	KI22_STRPU	P48872 strongyloce
12	703	28.3	1066	1	KL61_DROME	P48863 drosophila
13	702.5	28.3	1038	1	KL17_MOUSE	Q99PW8 mus musculus
14	692	27.9	1226	1	KF4A_XENLA	Q91784 xenopus lae
15	676.5	27.2	1225	1	KF4A_CHICK	Q90640 gallus gall
16	664.5	26.8	1031	1	KINH_STRPU	P35978 strongyloce
17	660.5	26.6	975	1	KINH_DROME	P17210 drosophila
18	656	26.4	672	1	OSM3_CABEL	P48873 caenorhabdi
19	656	26.4	796	1	KF3C_RAT	O55165 rattus norv
20	655	26.4	796	1	KF3C_MOUSE	O35066 mus musculus
21	654.5	26.4	963	1	KINH_HUMAN	P33176 homo sapien
22	653.5	26.3	963	1	KINH_MOUSE	Q61768 mus musculus
23	652.5	26.3	1749	1	KI3A_MOUSE	O9eqw7 mus musculus
24	652	26.3	793	1	KF3C_HUMAN	O14782 homo sapien
25	652	26.3	1805	1	KI3A_HUMAN	Q9nlh9 homo sapien
26	650	26.2	1826	1	KI3B_HUMAN	Q9nqt8 homo sapien
27	642.5	25.9	967	1	KINH_LOLPE	P21613 loligo peal
28	640	25.8	957	1	KF5C_HUMAN	O60282 homo sapien
29	638.5	25.7	1032	1	KINN_HUMAN	Q12840 homo sapien
30	636.5	25.6	956	1	KF5C_MOUSE	P28738 mus musculus
31	635.5	25.6	1027	1	KINN_MOUSE	P33175 mus musculus
32	633	25.5	784	1	KL68_DROME	P48867 drosophila
33	632.5	25.5	1057	1	KF11_HUMAN	P52732 homo sapien

34	630.5	25.4	1067	1	EG52_XENLA	Q91783 xenopus lae
35	629	25.3	1060	1	EG51_XENLA	P28025 xenopus lae
36	628	25.3	815	1	KINH_CABEL	P34540 caenorhabdi
37	623.5	25.1	554	1	KLP3_SCHPO	Q9us60 schizosacch
38	622	25.1	1584	1	U104_CABEL	P23678 caenorhabdi
39	621.5	25.0	935	1	KINH_SYNRA	O43093 syncephalas
40	611.5	24.6	1111	1	KIPI_YEAST	P28742 saccharomyc
41	610	24.6	928	1	KINH_NEUCR	P48467 neurospora
42	607	24.4	1056	1	KI25_ARATH	P82266 arabidopsis
43	603.5	24.3	1648	1	KF14_HUMAN	Q15058 homo sapien
44	600	24.2	1184	1	BIMC_EMENI	P17120 emericeella
45	599.5	24.1	1690	1	KF1A_HUMAN	Q12756 homo sapien

ALIGNMENTS

RESULT 1
CENE_HUMAN STANDARD; PRT; 2663 AA.
AC Q02224;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis.";
RL Nature 359:536-539 (1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT microtubule motor.";
RL EMBO J. 14:918-926 (1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63 (1998).
RN [4]
RP FARNESYLATION.
RX MEDLINE=20459117; PubMed=10852915;
RA Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L.,
RA Bishop W.R., Kirschmeier P.;
RT "Farnesyl transferase inhibitors block the farnesylation of CENP-E
RT and CENP-F and alter the association of CENP-E with the
RT microtubules.";
RL J. Biol. Chem. 275:30451-30457 (2000).
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Z15005; CAA78727.1; -

PIR; S28261; S28261.

HSSP; P17119; 3KAR.

Genew; HGNC:1856; CENPE.

DR GK; Q02224; -

DR MIM; 117143; -

DR GO; GO:0005699; C:kinetochore; TAS.

DR GO; GO:0005634; C:nucleus; TAS.

DR GO; GO:0008350; P:kinetochore motor activity; TAS.

DR GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.

DR GO; GO:0007080; P:mitotic chromosome movement; TAS.

DR InterPro; IPR001752; kinesin_motor.

DR Pfam; PF00225; kinesin; 1.

DR PRINTS; PR00380; KINESINHEAVY.

DR SMART; SM00129; KISC; 1.

DR PROSITE; PS00411; KINESIN MOTOR DOMAIN; 1.

DR PROSITE; PS00067; KINESIN MOTOR DOMAIN; 1.

KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;

Cell cycle; Centromere; Lipoprotein; Prenylation.

FT DOMAIN 1 335 KINESIN-MOTOR.

FT DOMAIN 336 2471 COILED COIL (POTENTIAL).

FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).

FT NP BIND 86 93 ATP (BY SIMILARITY).

FT LPID 2660 2660 S-farnesyl cysteine.

SEQUENCE 2663 AA; 312087 MW; CEFC13880C8C8CB8 CRC64;

Query Match 100.0%; Score 2483; DB 1; Length 2663;

Best Local Similarity 100.0%; Pred. No. 3.1e-136;

Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEEGAVAVCVVRPLNSREESLGETAQVYKNTDNNVYQVDSKSFNDRFVHGNETTKN 60

Db 2 AEEGAVAVCVVRPLNSREESLGETAQVYKNTDNNVYQVDSKSFNDRFVHGNETTKN 61

QY 61 VYEEIAPIIDSAIQGYNGTTFAYGQTASGTYTMGSEDLGVIPRAIHDFOKIKKFP 120

Db 62 VYEEIAPIIDSAIQGYNGTTFAYGQTASGTYTMGSEDLGVIPRAIHDFOKIKKFP 121

QY 121 DRELLRVSYMEIYNETITDLCTGQKMKPLIREDVNRNYYVADLTVEEVYVTSMAK 180

Db 122 DRELLRVSYMEIYNETITDLCTGQKMKPLIREDVNRNYYVADLTVEEVYVTSMAK 181

QY 181 ITKGEKSRHYGETKMNQSRSSHTIFRMILESRKGFPSNCEGSKVYSHLNLVDLAG 240

Db 182 ITKGEKSRHYGETKMNQSRSSHTIFRMILESRKGFPSNCEGSKVYSHLNLVDLAG 241

QY 241 AAGTGAAGVRLEKGCNINRSFILGQVTKKLSGQGVGFNYRDSKLTIRLONSLGNPK 300

Db 242 AAGTGAAGVRLEKGCNINRSFILGQVTKKLSGQGVGFNYRDSKLTIRLONSLGNPK 301

QY 301 TRITCTITPVDFETLTALQASTAKYWKNTPYNEVSTDEALLKRYRKEIMDLKQLEE 360

Db 302 TRITCTITPVDFETLTALQASTAKYWKNTPYNEVSTDEALLKRYRKEIMDLKQLEE 361

QY 361 VSLETRAQAMEKQDLAQLLEKDLQVNEKIEKNTLRLMTVTSLSLTQQLKAKRKRRV 420

Db 362 VSLETRAQAMEKQDLAQLLEKDLQVNEKIEKNTLRLMTVTSLSLTQQLKAKRKRRV 421

QY 421 TWCLGKINMKNSYADQFNIPNTITTKLISNLLREIDSVCSDFVNTLDTLSE 480

Db 422 TWCLGKINMKNSYADQFNIPNTITTKLISNLLREIDSVCSDFVNTLDTLSE 481

QY 481 LEWNPAT 487

Db 482 LEWNPAT 488

RESULT 2

KF3B HUMAN

ID KF3B HUMAN STANDARD; PRT; 747 AA.

AC Q15066;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Kinesin-like protein KIF3B (Microtubule plus end-directed kinesin motor 3B) (HH0048).

DE KIF3B OR KIAA0359.

GN KIF3B OR KIAA0359.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Primates; Catarrhini; Hominidae; Homo.

ON NCBI_TaxID=9606;

OX [1]

RP SEQUENCE FROM N.A.

TX TISSUE=Brain;

RX MEDLINE=97349984; PubMed=9205841;

RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro."

RL DNA Res. 4:141-150(1997).

[2]

RP SEQUENCE FROM N.A.

RX MEDLINE=21638749; PubMed=11780052;

RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R., Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L., Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Clee C.M., Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M., Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R., Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E., Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S., Leheslahti M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D., Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A., Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T., Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I., Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H., Rice C.D., Ross M.T., Scott C.E., Sehara H.K., Shownkeen R., Sims S., Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E., Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A., Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Whitehead S.B., Whittaker P., Willey D.B., Williams L., Williams S.A., Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S., Rogers J.;

RT "The DNA sequence and comparative analysis of human chromosome 20."

RL Nature 414:865-871(2001).

[3]

RP IDENTIFICATION IN A COMPLEX WITH SMC3 AND KIFAP3B.

RX MEDLINE=98175913; PubMed=9506951;

RA Shimizu K., Shiratani H., Honda T., Minami S., Takai Y.;

RT "Complex formation of SNAP/KAP3, a KIF3A/B ATPase motor-associated protein, with a human chromosome-associated polypeptide."

RL J. Biol. Chem. 273:6591-6594(1998).

CC -!- FUNCTION: Involved in tethering the chromosomes to the spindle pole and in chromosome movement. Microtubule-based anterograde translocator for membranous organelles. Plus end-directed microtubule sliding activity in vitro (By similarity).

CC -!- SUBUNIT: Heterodimer of KIF3A and KIF3B (By similarity). Interacts with the SMC3 subunit of the cohesin complex.

CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN II SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its


```

RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA MEDLINE=2388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S.C., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[3]
RP SEQUENCE FROM N.A.
RA Connolly K.S., Gunning K.M., Davis C.A., Kadner K., Subramanian S.,
RA Miguel T., Lewis K.D., Fridlyand J., Alcivare D., Benke J.A.,
RA Bondoc M., Bowen E., Chiang A., Critz P., Jaklevic M.A., Lindo K.,
RA Lindquist K., Miller C., Patel S., Pisciotta C., Riley B.E., Rojeski H.,
RA Sarmiento R., Yu C., Montenegro M., Aerts A., Chung A., Abramo A.,
RA Baker M., Gau C., Jett J., Ko C., Beall K., Woolley J.P., Stultz J.L.,
RA Kimmerly W., Martin C.H.;
RT "Sequencing of human chromosome 5";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MICROTUBULE-BASED ANTROGRADE TRANSLATOR FOR
CC MEMBRANOUS ORGANELLES. PLUS END-DIRECTED MICROTUBULE SLIDING
CC ACTIVITY IN VITRO.
CC -!- SUBUNIT: HETERODIMER OF KIF3A AND KIF3B.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC II SUBFAMILY.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF041853; AAC72294.1; --
DR EMBL; BC045542; AAH45542.1; --
DR EMBL; AC004237; AAC04475.1; ALT_SEQ.
DR HSSP; P17119; 3KAR.
DR Genew; HGNC:6319; KIF3A.
DR MIM; 604683; --
DR GO; GO:0006996; P:organelle organization and biogenesis; TAS.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
KW Motor protein; Microtubule; ATP-binding; Coiled coil; Neurone.
FT DOMAIN 1 350 KINESIN-MOTOR.
FT DOMAIN 351 599 COILED COIL (BY SIMILARITY).
FT DOMAIN 600 702 GLOBULAR.
FT NP_BIND 100 107 ATP (BY SIMILARITY).
FT DOMAIN 443 446 POLY-GLU.
FT CONFLICT 151 151 E -> G (IN REF. 2).
FT CONFLICT 170 170 E -> A (IN REF. 1).
FT CONFLICT 172 172 K -> I (IN REF. 2).
SQ SEQUENCE 702 AA; 80385 MW; 88555D21209B4E14 CRC64;
Query Match 29.1%; Score 723; DB 1; Length 702;
Best Local Similarity 35.8%; Pred. No. 8.5e-35;
Matches 193; Conservative 84; Mismatches 166; Indels 96; Gaps 14;
QY 6 VAVCVVRPLNSRERSL--GETAQVYVWKTDNNVIVQVDS-----KSFNDFRPHGNETTK 59
DB 15 VKVVVRCPPLNERKSCVQKQVSVDEMGTTTVHKTTSSNEPPKTFTEFTVFGPESKOL 74
QY 60 NYVEEIAPIIDSAIQNGYNTIFAYGQTASGTYTYMGSE---DHLGVIPRAIHDFQKI 116
DB 75 DYNLTARPIIDSVLEGYNTIFAYGQTGKTFTMEGVRAIPELRGIIPNSFAHIFGHI 134
QY 117 KFPF--DRFELLVSVMEYINETITDLCQTQOMKPLIIREDVNRNVYVADLTVEVVYTS 175
DB 135 AKAEGETRFLRVSVYLEYNEEVRDL--CKDQTORLEVKERPDVGYIKDLSAYVNNAD 193
QY 176 MALKWITKGEKSRHYGETKMNORSSRSHTIPRMILESREKGPSCGSKVYSHLNLVDL 235
DB 194 DMDRIMTGLHKNRVSGVATNMHSHSRSHAFTITIECKEKIDGNWH--VRMGKLUHLVDL 251
QY 236 AGSERAAGTGAAGVRLKEGCNINRSLFIIQGVKKLSGQGVGFYNYRDSKLTIRIQLNSL 295
DB 252 AGSERQAKTGATGQRLKEATKINLSLSTLGNVISALVDGK--STHVPYNSKLTIRIQLNSL 310
QY 296 GGNPKTRIICTITPV--SFDETLALQPASTAKYMNTPYNEVSTDEALLKRYRKEIMD 353
DB 311 GGNSTWMCANIGPADYNYDETISTLYRANAKNIKRNARINEDPKD--ALLRFQKEITEE 369
QY 354 LKKOLEE----- 360
DB 370 LKKLEEGEEISGDISSEDDDEGEVGEDEKRRKRRDQTGKKKVSPPDMIMQAKI 429
QY 361 ----VSLETRAQAMEKDLQALLE----EKOLQKVQNEKIENLRMLVTSSSLTLOQEL 412
DB 430 DEERKALETKLDMEERNEERKARAELEKREKOLLKAQAEHQ-----SLEKL 475
QY 413 KAKRRRVTWCLGINKMKNSYADQFNIPNI--TTTKHKLSINLLRIDSVCSESDV 470
DB 476 SALEKKVI---VGVVDLLAKAEQEKLEESNMELERRKRAEQLRRELEKQERLDI 531
RESULT 6
K121_STRPU STANDARD; PRT; 742 AA.
AC P46871;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Kinesin-II 95 kDa subunit (KRP-85/95 95 kDa subunit).
GN KRP95.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Egg;
RX MEDLINE=94050179; PubMed=8232586;
RA Cole D.G., Chinn S.W., Wedaman K.P., Hall K., Vuong T.,
RA Scholey J.M.;
RT "Novel heterotrimeric kinesin-related protein purified from sea
RL urchin eggs.";
RL Nature 366:268-270(1993).
CC -!- SUBUNIT: Heterotrimer of a 115 kDa subunit (KAP115) and two
CC kinesin-like subunits of 95 kDa (KRP95) and 85 kDa (KRP85).
CC -!- PTM: The N-terminus is blocked.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC II SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U00996; AAA87393.1; --
CC HSPSP; P17119; 3KAR.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINSEINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
KW Motor protein; Microtubule; ATP-binding; Coiled coil.
FT DOMAIN 1 337 KINESIN-MOTOR (BY SIMILARITY).
FT DOMAIN 338 613 COILED COIL (BY SIMILARITY).
FT DOMAIN 614 742 GLOBULAR (BY SIMILARITY).
FT NP BIND 95 102 ATP (POTENTIAL).
SQ SEQUENCE 742 AA; 84202 MW; 47C40A367BAA77B5 CRC64;

Query Match 29.1%; Score 722; DB 1; Length 742;
Best Local Similarity 37.6%; Pred. No. 1e-34;
Matches 198; Conservative 74; Mismatches 191; Indels 64; Gaps 13;

QY 6 VAVCVVRPLNSRESLGESLGETAQVYKTDNNVIYQV-----DGSKSFNDRVPHGNETT 58
DB 9 VKVYVRCRPLNGKEKADGRSRIYDMDVDAGQVVRNPKADASEPDKAFTDQVYDMNCQ 68
QY 59 KNYVEEIAAPIIDSAIQGYNGTIFAYGQTASGKTYTMMG---SEDLGLVPIRAIHDFQK 115
DB 69 IDLYDETRSLVESVLQGFNGTIFAYGQTGSGKTYTMMG---SEDLGLVPIRAIHDFQK 128
QY 116 IKKPPDREFLLRVSYMEIYNETITDLCCTQKPKLIIREDVNRNRYVADLVEEYVY 175
DB 129 IARTQNOQFLVRASYLEIYQBEIRDLLAKQOK-KRLDLKERPDGTGVYKOLSSFTVSKV 187
QY 176 MALXWITGKESRHYGETKMNQSRSHITPRMILESEKCEPNSCEGSKVSHNLVDL 235
DB 188 EIEHVMVTGNNRSVSGTNMHEHSRSHAFITIECSLGVG--ENHIVGKLNLDL 245
QY 236 AGSERRAQTGAAGVRLKEGNCINRSLFTLGIVIKLSDGVGGFYNRDSKILRLQNSL 295
DB 246 AGSERQATGATGDLKATKINLSALGNVISALVDGK--SHIYEDSKILRLQNSL 304
QY 296 GGNPKTRIICTIPVS--FDETLALQFASTAKYKNTPTPYNEVSTDBALLKRYKREIMD 353
DB 305 GGNKATVMVANMGFASYNFDETTITLRYANRAKNIKPKINEDPKD--ALLREFFQEEISR 363
QY 354 LKKQL-----EEVSLTRAQAMEKQDLAOLLEKD 383
DB 364 LKQALDKKPSDGRKKKRPGEQGGDDIEDTEEGEDMEDEEMVYKESQKLEBEKE 423
QY 384 LLQVQNEKIEMLTSLTSLTQBELKAKRRVTVTCGLKINKMKNKNYAQFNPT 443
DB 424 KIMANQSMIAEBEKQL--SEVQKQGEIK-KEHQKEMLEKGIKAMESKLLVGKGSIVD 480
QY 444 NITTKHLSINLL-----REIP-ESVCSSESVFNTDLTSLIE 482
DB 481 HTNEQQRKIEBQRLLAEBEKNRERDMERLKEQD-----DKTVEIE 521

RESULT 7
ID FL10 CHLRE STANDARD; PRT; 786 AA.
AC P46869;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Kinesin-like protein FLA10 (KIF1 protein).
GN FLA10.
OS Chlamydomonas reinhardtii.

OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J37;
RX MEDLINE=94299638; PubMed=8027176;
RA Walther Z., Vaehishtha M., Hall J.L.;
RT "The Chlamydomonas FLA10 gene encodes a novel kinesin-homologous
protein";
RL J. Cell Biol. 126:175-188(1994).
CC -!- FUNCTION: Probably involved in flagellar assembly and maintenance.
CC May play a role in flagellar synthesis.
CC -!- TISSUE SPECIFICITY: Flagellar axoneme.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC II SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; L33697; AAA21738.1; --
DR PIR; A53939; A53939.
DR HSPSP; P17119; 3KAR.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINSEINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
KW Motor protein; Microtubule; ATP-binding; Coiled coil.
FT DOMAIN 1 358 KINESIN-MOTOR (BY SIMILARITY).
FT DOMAIN 367 687 COILED COIL (POTENTIAL).
FT DOMAIN 688 786 GLOBULAR (POTENTIAL).
FT NP BIND 97 104 ATP (POTENTIAL).
FT DOMAIN 388 391 POLY-GLY.
FT DOMAIN 705 714 POLY-GLY.
FT DOMAIN 756 759 POLY-ASP.
SQ SEQUENCE 786 AA; 86671 MW; F90969203EB79F1B CRC64;

Query Match 28.8%; Score 716; DB 1; Length 786;
Best Local Similarity 36.3%; Pred. No. 2.5e-34;
Matches 194; Conservative 86; Mismatches 176; Indels 78; Gaps 16;

QY 5 AVAVCVVRPLNSRESLGESLGETAQVYKTDNNVIY-----QVDGS---KSFNDRVPHGNET 57
DB 10 SVKVVVRCRPLNGKEKADGRSRIYDMDVDAGQVVRNPKADASEPDKAFTDQVYDMNCQ 69
QY 58 KNYVEEIAAPIIDSAIQGYNGTIFAYGQTASGKTYTMMGSEDH---LGVIPIRAIHDFQ 114
DB 70 QRVDFDITARLIDSCLEGNGTIFAYGQTGSKSHMEGKDEPELRLGIPNTFRVYFE 129
QY 115 KI-KKFPDREFLLRVSYMEIYNETITDLCCTQKPKLIIREDVNRNRYVADLVEEYVY 173
DB 130 IIAARDSGTKEFLVRSSYLEIYNEEVROLL-GKDSHKKMELKESPDGQVYVVDLSQVCKN 188
QY 174 SEMALKWITGKESRHYGETKMNQSRSHITPRMILESEK-----GEPSNCEGS 224
DB 189 YEEMNKVLLAGKDRNRQVGTALNQDSRSRSHIFITIECIKLSAAAKQKAGKDDSNH 248
QY 225 VKVSHNLVDLAGSERRAQTGAAGVRLKEGNCINRSLFTLGIVIKLSDGVGGFYNRDSK 284
DB 249 VRVGKLNLDLAGSERQDKTGTDLKKEGINKINLSLALGNVISALVDGK--SGHIPYRD 307
QY 285 SKLTRLQNSLGGNPKTRIICTIPV--SFDETLALQFASTAKYKNTPTPYNEVSTDBEA 342
DB 308 SKLTRLQNSLGGNPKTRIICTIPV--SFDETLALQFASTAKYKNTPTPYNEVSTDBEA 366
QY 343 LLKRYKREIMDLKQL-----EEVSLTRAQAMEKQD 374

Db 367 MLRQFOEIKKLEQAARAAGGGPITMPSGGSPQKXIVERTEEDPIDA---IKAQ 423
 QY 375 LAQLLEK-----DLQKVQNE---KINLRLMTLVTSSSLTQELKA-----KKRR 419
 Db 424 MRLEAKMKSDISTALDKAREEAAAKKQLQAIIDQCKTEAOKKAARDALKQAEE 483
 QY 420 VTWCLGKINKMKNVADQNFITNTKTHKL---SINLLREIDSVCSSESV 470
 Db 484 ARAIAGAIEKEQ-----EKAVLESRIKMEGKIVGGVNMLEKVDLQKQSEDI 533

RESULT 8

KF4A_HUMAN
 ID KF4A_HUMAN STANDARD; PRT; 1232 AA.
 AC Q95239; Q9NNY6; Q9NY24; Q9UMW3;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 16-OCT-2001 (Rel. 40, last sequence update)
 DT 28-FEB-2003 (Rel. 41, last annotation update)
 DE Chromosome-associated kinesin KIF4A (Chromokinesin).
 GN KIF4A OR KIF4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymphocytes;
 RA Villard L.
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RX MEDLINE=20435301; PubMed=10978527;
 RA Oh S.J., Hahn H., Torrey T.A., Shin H., Choi W., Lee Y.M.,
 RA Morse H.C. III, Kim W.;
 RT "Identification of the human homologue of mouse KIF4, a kinesin
 superfamily motor protein";
 RL Biochim. Biophys. Acta 1493:219-224(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Retinoblastoma;
 RA Rentsch A., Neumann T., Rommerskirch W.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 128-1232 FROM N.A.
 RC TISSUE=Retinoblastoma;
 RX MEDLINE=97911419; PubMed=9168136;
 RA Yan R.-T., Wang S.-Z.;
 RL Gene 189:263-267(1997).
 CC "Increased chromokinesin immunoreactivity in retinoblastoma cells";
 CC "FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR
 SPINDLE STABILIZATION (BY SIMILARITY).
 CC "SUBCELLULAR LOCATION: Nuclear. Associated with mitotic
 chromosomes (by similarity).
 CC "TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEMATOPOIETIC TISSUES,
 FETAL LIVER, SPLEEN, THYMUS AND ADULT THYMUS AND BONE MARROW.
 CC LOWER LEVELS ARE FOUND IN HEART, TESTIS, KIDNEY, COLON AND LUNG.
 CC "SIMILARITY: Belongs to the kinesin-like protein family.
 CC Chromokinesin subfamily.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sb.ch).

 DR EMBL; AF179308; AAD51855.1; -
 DR EMBL; AF071592; AAD05492.2; -
 DR EMBL; AJ271784; CAB75427.1; -
 DR EMBL; AF277375; AAF86334.1; -
 DR HSSP; P17119; 3KAR.

Genew; HGNC:13339; KIF4A.
 GO: GO:0005737; C:cytoplasm; TAS.
 GO: GO:0005876; C:spindle microtubule; TAS.
 GO: GO:0003777; F:microtubule motor activity; TAS.
 GO: GO:0008089; P:anterograde axon cargo transport; TAS.
 GO: GO:0006996; P:organelle organization and biogenesis; TAS.
 InterPro; IPR001752; kinesin_motor.
 Pfam; PF00225; kinesin; 1.
 PRINTS; PR00380; KINESINHEAVY.
 SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; DNA-binding;
 KW Nuclear protein; Coiled coil.
 FT DOMAIN 1 349 KINESIN-MOTOR.
 FT DOMAIN 350 999 COILED COIL (BY SIMILARITY).
 FT BIND 1000 1232 GLOBULAR.
 FT NP_BIND 88 95 ATP (POTENTIAL).
 FT CONFLICT 223 223 R -> G (IN REF. 2).
 FT CONFLICT 231 231 S -> T (IN REF. 4).
 FT CONFLICT 286 286 V -> A (IN REF. 2).
 FT CONFLICT 422 422 L -> W (IN REF. 2).
 FT CONFLICT 564 564 L -> H (IN REF. 4).
 FT CONFLICT 564 564 L -> P (IN REF. 2).
 FT CONFLICT 600 600 K -> E (IN REF. 3).
 FT CONFLICT 668 668 R -> K (IN REF. 3 AND 4).
 FT CONFLICT 928 928 Q -> P (IN REF. 1).
 FT CONFLICT 958 958 Q -> R (IN REF. 3).
 FT CONFLICT 960 960 L -> Q (IN REF. 1).
 FT CONFLICT 996 997 LL -> S (IN REF. 4).
 FT CONFLICT 1003 1014 OKHLPKDTLLSP -> RYLPRIPFFYLQ (IN REF. 4).
 FT CONFLICT 1022 1022 P -> Q (IN REF. 2).
 FT CONFLICT 1077 1077 K -> N (IN REF. 2).
 FT CONFLICT 1138 1138 G -> S (IN REF. 2).
 SQ SEQUENCE 1232 AA; 139908 MW; FF74052A17A8E8F7 CRC64;
 Query Match 28.7%; Score 713.5; DB 1; Length 1232;
 Best Local Similarity 35.1%; Pred. No. 6.3e-34;
 Matches 182; Conservative 86; Mismatches 191; Indels 59; Gaps 11;
 QY 6 VAVCVVRPLNSRBSLG-ETAAQVYWKTDNNVIYQVDSKSTNFDVRPHGNETTKNVEE 64
 Db 10 VRVALRCPLVPEKEISEGCMCLSFVGPQVVGTD--KSTYDFVDFPSTEQBEVFT 67
 QY 65 TAAPIDSALQYNGTIFAYGQTASGKYTMG-----SEHLGVIPRAIHDFQKIK 117
 Db 68 AVAPLKGKGYNATVLAYGQTSKTYSMGAYTAEQENETVGVIPRVQLLFKEID 127
 QY 118 KFPDREFLLRVSYMEIYNETTDLCCGQKPKPLIREDVNVNVDLTERVWYSEMA 177
 Db 128 KKSDEFILKVSYLEIYNEEILDLCPREKQAINREDPKGIKIVGLTEKTVLVALDT 187
 QY 178 LKWITGKSRHYGTEKMNORSRSHITFRMILESREKGEPSNCGSVKSHNLVLDLAG 237
 Db 188 VSCLEQGNNSRTVASTAMNSQSSRSHAFITSLQEKSD---KNSSFRSKLHLVDLAG 243
 QY 238 SERAAQTGAAGVRLKEGNCINRSLPILQVTKLSDGVGVGFYINRDSKLTILQNSLGG 297
 Db 244 SERQKTKAEGDRLKEGINRGLICLGNVISALGDDKGGFVPYRDSKLTLLQDSLGG 303
 QY 298 NPKTRIICTINPV--SPDETALQFASTAKYMKNTPVNVESTDALLKRYKREIMDLK 355
 Db 304 NSHTLMIAVCSPADSNDEETLNTURYADRARKINKKPIV-----IDPQTAELNHLK 355
 QY 356 KQLBEVSL-----ETRAQAMEKDLQALKEKDLQKLVQNEKIENL 396
 Db 356 QVQQLQVLLLAQHGGTLPGSITVEPSNLSLMEKNQ--SLIVENEKLRLGLSFAAQGT 413
 QY 397 TRML---VTSSSLTLQOELKAKRKRVTCWLGKINK-----MKNSNYADQFNPTNTTKT 449
 Db 414 AQMLERIILTEQANEKKNNAKLEELRQHAACKLDLQKLVTETLEDQELKENVEIICNLQOLI 473
 QY 450 HKLSINLLREIDSVCSSESVFNSLTDLTSLSEIENPAT 487

QY 349 KEIMDLKKOLEVSLLETRAQAMEKDKQLAQLL-----BEKOLLQKV----- 388
 Db 355 ERIKKLKAIL-----TQMGSPSLLSALLSQVPPDPVQVEKLLPQVIOHDMEAEX 406
 QY 389 ---ONEKIENTLRLMTVSSSLTIQELKAKRKRVTWCLGKINKNSNVADOFNPTNI 445
 Db 407 QLIREFEYERLARL---KADYKAEQSRARLEEDIT-----AMRNSYDVRSLTLEENL 456
 QY 446 TTTKHLSINLLREIDESVCSDFNSLTDLTSLSEIWNPA 486
 Db 457 RKETEAV-----LQGVLYKAEVMSRA-EFASSAEYPPA 489

RESULT 10
 KF4A MOUSE
 ID KF4A MOUSE STANDARD; PRT; 1231 AA.
 AC P33174;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chromosome-associated kinesin KIF4A (Chromokinesin).
 GN KIF4A OR KIF4 OR KNS4.
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
 RC STRAIN=ICR; TISSUE=Brain;
 RX MEDLINE=95014709; PubMed=7929562;
 RA Sekine Y., Okada Y., Noda Y., Kondo S., Aizawa H., Takemura R., Hirokawa N.;
 RT "A novel microtubule-based motor protein (KIF4) for organelle transports, whose expression is regulated developmentally.";
 RL J. Cell Biol. 127:187-201(1994).
 RN [2]
 RP SEQUENCE OF 91-240 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93077686; PubMed=1447303;
 RA Aizawa H., Sekine Y., Takemura R., Zhang Z., Nangaku M., Hirokawa N.;
 RT Kinesin family in murine central nervous system.";
 RL J. Cell Biol. 119:1287-1296(1992).
 CC -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOlar SPINDLE STABILIZATION.
 CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic chromosomes.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN PYRAMIDAL CELLS IN JUVENILE HIPPOCAMPUS, GRANULAR CELLS IN JUVENILE CEREBELLAR CORTEX AND IN ADULT SPLEEN.
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family. Chromokinesin subfamily.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

 EMBL; D12646; BAA02167.1; -.
 DR PIR; A54803; A54803.
 DR HSSP; P17119; 3KXR.
 DR MGP; MGI:108389; Kif4.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PRO0380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; DNA-binding;

KW Nuclear protein; Coiled coil.
 FT DOMAIN 1 350 KINESIN-MOTOR.
 FT DOMAIN 351 1000 COILED COIL (BY SIMILARITY).
 FT DOMAIN 1001 1231 GLOBULAR.
 FT NP BIND 88 95 ATP (POTENTIAL).
 FT CONFLICT 112 112 I -> S (IN REF. 2).
 SQ SEQUENCE 1231 AA; 139551 MW; F34F2C2D21158FE4 CRC64;

Query Match 28.6%; Score 710.5; DB 1; Length 1231;
 Best Local Similarity 35.8%; Pred. No. 9.5e-34;
 Matches 186; Conservative 81; Mismatches 184; Indels 69; Gaps 14;

QY 6 VAVCVRVRLNLSREESLG-ETAOVYWKTDNNVIYQVDSKSNFDFVHGNETHNKNYEE 64
 Db 10 VRVALRCRPLVSKKEIKGQCOTCLSFVPGPVV--VGNDKSFYDFVDFPSTEEQEEVNT 67
 QY 65 IAAPIIDSAIQYNGTIFAYGOTASGKTYTMGS---EDH---LGVIRAHIDFQKIK 117
 Db 68 AVAPLIKGVFKGNATVLAAYGOTSGKTYSMGAYTAQEHSAGIIVFVIOQLLKEIN 127
 QY 118 KFPDFEFLRVSYMEIYNETITDLCGT-OKMKPLIREDVARNVYVADLTTEVVYVTS 176
 Db 128 KKSDFEFTLKVSYLEIYNEEILDLCCSREKATQINIREDPKEGIVGLTEKTVLVASD 187
 QY 177 ALKWITKGEKSHYGETKMNQSSSHTIFRMLESREKGEPSNCEGVSVKVSHNLVDLA 236
 Db 188 TVSCLEQGNNSRTVASTAMNSOSSRSHAFTISIEQRKK---NDKNSFRSKLHLVDLA 243
 QY 237 GSERAAQTGAAGVRLKEGNCINRSLFIQVQIKKLSGQGVGFVYRDSKLTILQNSLG 296
 Db 244 GSERQKTKAEGDRLREGINRGLLGLNGLSALGDDCKGNFVYRDSKLTILQDLSLG 303
 QY 297 GNPKTRITCTTPV--SPDETTLAQFSTAKYMNTPVYNEVSTDEALLKYRKEIMDL 354
 Db 304 GNSHTLMIACTVSPADSNLEETLTLRYADRAKIKNKPIN-----IDPQAAELNHL 355
 QY 355 KKOLEEVSL-----ETRAQAMEKDKQLAQLLBEKOLLQKVQNEKIEIN 395
 Db 356 KQVQOQLILLQAHGGLPGDINVEPSENLQSLMEKNQ--SLVRENKLSRGLSEAAQ 413
 QY 396 ITRML---VTSSSLTLOELKAKRRVTCWLGKINK---MKNSNYADQFNIPNTITK 448
 Db 414 TAQMLERILTEQANEKKNKLEELRRHACKVDIQLVTELDDELKENIELICNLQOV 473
 QY 449 THKLS-----INLLREIDESVCS-----ESDVFS 472
 Db 474 IATLSDEAAACWTATIDTAGEADTQVQSSPDTSRSSDVFS 513

RESULT 11
 KI22_STRPU
 ID KI22_STRPU STANDARD; PRT; 699 AA.
 AC P46872;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Kinesin-11 85 kDa subunit (KRP-85/95 85 kDa subunit).
 GN KRP85.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
 CC Strongylocentrotus.
 OX NCBI_TaxID=7668;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Egg;
 RX MEDLINE=94050179; PubMed=8232586;
 RA Cole D.G., Chinn S.W., Wedaman K.P., Hall K., Vuong T., Scholey J.M.;
 RT "Novel heterotrimeric kinesin-related protein purified from sea urchin eggs.";
 RL Nature 366:268-270(1993).
 CC -!- SUBUNIT: Heterotrimer of a 115 kDa subunit (KAP115) and two

CC kinesin-like subunits of 95 kDa (KRP95) and 85 kDa (KRP85).
 CC -!- BTM: The N-terminus is blocked.
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
 CC II SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L16993; AAA16098.1; --
 CC PIR: S38982; S38982.
 CC HSP: P17119; 3KAR.
 CC InterPro: IPR001752; kinesin_motor.
 CC Pfam: PF00225; kinesin; 1.
 CC PRINTS: PR00380; KINESINHEAVY.
 CC SMART: SM00129; KISC; 1.
 CC PROSITE: PS00411; KINESIN MOTOR DOMAIN1; 1.
 CC PROSITE: PS00667; KINESIN MOTOR DOMAIN2; 1.
 CC Motor protein; Microtubule; ATP-binding; Coiled coil.
 CC DOMAIN 1 340 KINESIN-MOTOR (BY SIMILARITY).
 CC COILED COIL (BY SIMILARITY).
 CC DOMAIN 341 619
 CC COILED COIL (BY SIMILARITY).
 CC DOMAIN 620 699 GLOBULAR (BY SIMILARITY).
 CC NP_BIND 97 104 ATP (POTENTIAL).
 CC SEQUENCE 699 AA; 78697 MW; 783866111CB08190 CRC64;
 CC -----
 CC Query Match 28.4%; Score 704.5; DB 1; Length 699;
 CC Best Local Similarity 36.2%; Pred. No. 1e-33;
 CC Matches 200; Conservative 76; Mismatches 152; Indels 125; Gaps 19;
 CC -----
 CC QY 6 VAVCVVRPLNSRESLG-----ETAOVYWKTDNNVIVQDGSKSFNDRVPHG 54
 CC DB 11 VRVVRVCRPLNSKETGQFKSVKXMDMGRGVQV---TNPNA-PSGEPKSFETDTVPAP 66
 CC QY 55 NETTKNYEETAAPITDAISAOYNGTIFAYGOTASGKTYTWG--SEDLH-GVIPRAIHD 111
 CC DB 67 GAKQTDVNVQARPLVDIAIEGNGTIFAYGOTGTGKFTWEGVRSQPELAGIIPNSPAH 126
 CC QY 112 IFQKIKKPPDR-EFLLRVSYMEIYNETITDLCGQTKQKMLIREDVNRNVYVADLTVEV 170
 CC DB 127 IFGHIAKQENVRFLVRVSYLEIYNEEVKDLL-GRDQQRLEVKRDPGVGVVVDLSAFV 185
 CC QY 171 VYTEMALKWITKESKSHYGETKQNRSSRSHITIFRMILLESREKGEPSNCEGSKVSHL 230
 CC DB 186 VNNADMDRIMTGLNKNRSGVATNNSSSRSHAFITTLERSDMG--LDKEQHVVRGKL 243
 CC QY 231 NIVDLAGSERAQTAAGVRLKEGNCINRSLFILQVTKKLSDGQVGGFINVRDSKLTRI 290
 CC DB 244 HMVDLAGSERQTKTQATGRLKEATKINLSLTLGNVLSLVLDGK-STHPIVRNSKLTREL 302
 CC QY 291 LQNSLGGNPKTRIICITIPV--SPDETALQFASAKYMKNTPTVNEVSTDEALLKRYR 348
 CC DB 303 LQDSLGGNAKTVWCANIGPAENYDETISTLYANRAKNIKNAKINEDPKD-ALLREFFQ 361
 CC QY 349 KEIMDLKKQLE-----VSLETRAQAM 370
 CC DB 362 KEIEELKQKISSGEGLDDEESGSEGEAGEGGVKKKPKKPKKPKKSPLEMA-AM 420
 CC QY 371 EK--DQLAQLEE-KOLL-----QKVNQKIEKNTLRLVLT- 402
 CC DB 421 QKKIDEEKALKREKMDVREDNTVHRELORRESELHKAQDDQKIENKLNIAIQKLLIVG 480
 CC QY 403 -----SSSLTIQQELKAKRVRVTCGLKINMKNSNVADQFNPTNIT 446
 CC DB 481 GVDLLAKSBEQQLLEQSALENKERMKAQESMR-----KNMEEREQERMDEEKYS 531
 CC QY 447 T-----KTHKL 452
 CC DB 532 SLQDEAHGKTCKL 544

RESULT 12
 KL61 DROME STANDARD; PRT; 1066 AA.
 AC P46863; Q8T0A6; Q9W018;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 GN Bipolar kinesin KRP-130 (Kinesin-like protein Klp61F).
 GN KLP61F OR KLP2 OR CG9191.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 CX NCBI_TaxID=7227;
 [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Embryo;
 RC MEDLINE=94043448; PubMed=8227131;
 RX Heck M.M.S., Pereira A., Pesavento P.A., Yannoni Y., Spradling A.C.,
 RA Goldstein L.S.B.;
 RT "The kinesin-like protein KLP61F is essential for mitosis in
 RL Drosophila.";
 RL J. Cell Biol. 123:665-679 (1993).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Berkley;
 RC MEDLINE=20196006; PubMed=10731132;
 RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.G., Rogers Y.-H.C., Blazer G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.B.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J.J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Strong R., Sun E.,
 RA Spier E., Spradling A.C., Stapleton M., Venter E., Wang A.H., Wang X.,
 RA Svirskaas R., Tector C., Turner R., Stupski M.P., Smith T.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195 (2000).
 [3]
 RN REVISIONS.
 RP MEDLINE=22426069; PubMed=12537572;
 RX Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.B.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,

Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
systematic review";
RN Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
Rubin G.M., Celniker S.E.;
RA Drosophila full-length cDNA resource";
RT "Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN [5]
RP SEQUENCE OF 228-357 FROM N.A.
RC STRAIN=DP CN BW;
RX MEDLINE=92020874; PubMed=1243206;
RA Stewart R.J., Pesavento P.A., Woerpel D.N., Goldstein L.S.B.;
RT "Identification and partial characterization of six members of the
kinesin superfamily in Drosophila";
RN Proc. Natl. Acad. Sci. U.S.A. 88:8470-8474(1991).
RP [6]
RP PARTIAL SEQUENCE, AND IDENTIFICATION AS KRP-130.
RC MEDLINE=97078747; PubMed=8918872;
RA Kashina A.S., Scholey J.M., Leszyk J.D., Saxton W.M.;
RT "An essential bipolar mitotic motor";
RN Nature 384:225-225(1996).
RP -!- FUNCTION: IMPORTANT ROLE IN MITOTIC DIVIDING CELLS. MICROTUBULE
MOTOR REQUIRED FOR SPINDLE BODY SEPARATION. SLOW PLUS-END DIRECTED
MICROTUBULE MOTOR CAPABLE OF CROSS-LINKING AND SLIDING APART
ANTIPARALLEL MICROTUBULES, THEREBY PUSHING APART THE ASSOCIATED
SPINDLE POLES DURING SPINDLE ASSEMBLY AND FUNCTION.
CC -!- SUBUNIT: Homotetramer.
CC -!- DEVELOPMENTAL STAGE: SPECIFICALLY EXPRESSED IN PROLIFERATING
TISSUES DURING EMBRYONIC AND LARVAL DEVELOPMENT.
CC -!- PTM: PHOSPHORYLATION DURING MITOSIS AT THR-933 CONTROLS THE
ASSOCIATION OF KLP61F WITH THE SPINDLE APPARATUS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BIRC
SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; U01842; AAA03718.1; --
DR EMBL; A8303471; AAF47458.2; --
DR EMBL; AY069442; AAL39587.1; --
DR EMBL; M74428; AAA28655.1; --
DR FIR; A48669; A48669.
DR HSSP; P47119; 3KAR.
DR FlyBase; FBgn0004378; Klp61F.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005871; C:kinesin complex; IDA.
DR GO; GO:0003774; F:motor activity; IDA.
DR GO; GO:0007100; P:centrosome separation; IGI.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin_1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KUSC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
KW Motor protein; Cell division; Microtubule; ATP-binding; Coiled coil;
KW Mitosis; Cell cycle; Phosphorylation.
FT DOMAIN 17 361 KINESIN-MOTOR (BY SIMILARITY).
FT DOMAIN 362 462 COILED COIL (POTENTIAL).
FT DOMAIN 540 569 COILED COIL (POTENTIAL).
FT DOMAIN 639 738 COILED COIL (POTENTIAL).

QY 411 ELKAKRKRVTWCLGKINKKNSNVADQ---FNIPNTIT-----TKTHKLSINLLREI 460
 Db 401 DTEAEKQLIREYERLARKADYEAQSRVQLQEDITAMRNSYDVKLSTLQENLRKEX 460
 QY 461 DESVCSSESDVFNLTDLTSEIE 482
 Db 461 ETEAILKAEVLCKT-EVMSRAE 481

RESULT 14
 KF4A_XENLA STANDARD; PRT; 1226 AA.
 ID Q91784; Q9PSI0;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chromosome-associated kinesin KLP1 (Chromokinesin).
 GN KLP1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 OC Xenopodinae; Xenopus.
 OC NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
 RC TISSUE=Oocyte;
 RX MEDLINE=95236444; PubMed=7720067;
 RA Vernos I., Raats J., Hirano T., Heasman J., Karsenti E., Wylie C.;
 RT "Xklp1, a chromosomal Xenopus kinesin-like protein essential for spindle organization and chromosome positioning.";
 RL Cell 81:117-127(1995).
 RN [2]
 RP SEQUENCE OF 9-338 FROM N.A.
 RX MEDLINE=93246065; PubMed=8482413;
 RA Vernos I., Heasman J., Wylie C.;
 RT "Multiple kinesin-like transcripts in Xenopus oocytes.";
 RL Dev. Biol. 157:232-239(1993)
 CC -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR SPINDLE STABILIZATION.
 CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic chromosomes.
 CC -!- TISSUE SPECIFICITY: Expressed in oocytes, eggs, testes and brain.
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family. Chromokinesin subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X82012; CAA57539.1; --
 PIR; I51617; I51617.
 DR HSP; P17119; 3KAR.
 DR InterPro; IPR001752; kinesin motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; DNA-binding;
 KW Nuclear protein; Coiled coil.
 FT DOMAIN 1 350 KINESIN-MOTOR.
 FT DOMAIN 351 1006 COILED COIL (BY SIMILARITY).
 FT DOMAIN 1007 1226 GLOBULAR.
 FT NP BIND 87 94 ATP (POTENTIAL).
 FT CONFLICT 163 163 I -> L (IN REF. 2).
 SQ SEQUENCE 1226 AA; 138923 MW; 7F0275FCF3316697 CRC64;

RX MEDLINE=20370907; PubMed=10846156;
 RA Setou M., Nakagawa T., Seog D.-H., Hirokawa N.;
 RT "Kinesin superfamily motor protein Kif17 and mLin-10 in NMDA receptor-containing vesicle transport.";
 RL Science 288:1796-1802(2000).
 CC -!- FUNCTION: Transports vesicles containing N-methyl-D-aspartate (NMDA) receptor 2B along microtubules.
 CC -!- SUBUNIT: Interacts with LIN-10 PDZ domain.
 CC -!- TISSUE SPECIFICITY: Neuronal-specific.
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AB008867; BAB21099.1; --
 DR HSP; P17119; 3KAR.
 DR MGD; MGI:1098229; Kif17.
 DR GO; GO:0005871; C:kinesin complex; IDA.
 DR GO; GO:0003777; F:microtubule motor activity; IDA.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0007017; P:microtubule-based process; IDA.
 DR GO; GO:0016132; P:vesicle-mediated transport; IDA.
 DR InterPro; IPR001752; P:vesicle-mediated transport; IDA.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; Coiled coil; Transport;
 KW Protein transport.
 FT DOMAIN 1 265 KINESIN-MOTOR.
 FT DOMAIN 346 470 COILED COIL (POTENTIAL).
 FT DOMAIN 748 855 COILED COIL (POTENTIAL).
 FT NP BIND 91 98 ATP (POTENTIAL).
 FT SEQUENCE 1038 AA; 116372 MW; 2BED852A3AFDBD46 CRC64;

Query Match 28.3%; Score 702.5; DB 1; Length 1038;
 Best Local Similarity 37.1%; Pred. No. 2.2e-33;
 Matches 186; Conservative 88; Mismatches 179; Indels 49; Gaps 15;

QY 5 AVAVCVRVPLNLSRESLG-ETFAQVYVWKTDDNNVIVQDGS-----KSFNFRVFGHNETT 58
 Db 5 SVKVVRCRPMKRRERELSCQSVTVVDSARGQCFTQNGAADBPCKQFTFGAYIEHFT 64
 QY 59 KNVYEIAPIAIDSATQYNGTIFAYGOTASGKTYTMGSED---HLGVIPRAIHDFQK 115
 Db 65 EQIYNEIAYPLVEGVTEGNTIFAYGOTGSGSKSFTMQGLPPPCQGIIPRAFEHVFS 124
 QY 116 IKKFPDFELRLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVYVADLTVEVVTSE 175
 Db 125 VQCAENTKELVRSAYLEIYNEDVHDL-LGADTQRLKELKHEPKGVYVKGSLMHTVHVA 183
 QY 176 MALKWTIKGEKGRHYGETKQKQSSRSHTIFRMILE---SREKGEPSNCEGSKVSHNL 232
 Db 184 QCBVMTGWKQRAVGYTLNKKDSSRSHTIFNIBIYAVDERG-----KDLHRAQKLN 238
 QY 233 VDLAGSRAAQTCAAGVRKKEGNCINERSLFILOVTKKLSGQGVGFYNRDLSKLTRLQ 292
 Db 239 VDLAGSRQKQTGATGRLKEATKINLSALGNVISALVDGRC-KHPIPYRDSKLTRLQ 297
 QY 293 NSLGGNPKTRIICTIPV--SPDETALQFASTAKYMKQTPPVNEVSTDEALLKRYKE 350
 Db 298 DSLGGNTKLMWACLSPADNNYDETILSTLYANRAKNIKKNPRINEDPKD-ALLREYQEE 356
 QY 351 IMDLKKOLEVSLTETAAQMEKQLAQLLEKDLLOKVQNEKLENLRLMLVTSSSLTLOQ 410
 Db 357 IKRLKAIL-----AQQMGPGNLSALLSTQTTPPGPVQSEE-----KLIL---SPTIVQ 400

us-10-045-631b-88_copy_2_488.rsp

Fri Aug 6 10:49:14 2004

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:27:31 ; Search time 20.6858 Seconds
(without alignments)
7428.155 Million cell updates/sec

Title: US-10-045-631b-88_COPY_2_488

Perfect score: 2483

Sequence: 1 AEGAVAVCVVRPLNSREE.....SDVFSNTLDTLSEIWNPAT 487

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2021.5	91.4	549	11 Q7TPX4	Q7tpx4 mus musculus
2	1584	63.8	2954	13 Q42263	Q42263 xenopus lae
3	880	35.4	807	10 Q94HV9	Q94hv9 arabidopsis
4	880	35.4	823	10 Q9S7P3	Q9s7p3 arabidopsis
5	856	34.5	459	10 Q9SS30	Q9ss30 arabidopsis
6	844.5	34.0	888	10 Q9L062	Q9l062 arabidopsis
7	801	32.3	1885	5 Q869B8	Q869b8 dictyosteli
8	780.5	31.4	2013	5 Q9VK10	Q9vk10 drosophila
9	780.5	31.4	2244	5 Q9NCGO	Q9ncgo drosophila
10	769.5	31.0	1055	10 Q8RWM4	Q8rwm4 arabidopsis
11	769.5	31.0	1055	10 Q8W5R5	Q8w5r5 arabidopsis
12	759	30.6	1459	3 P87198	P87198 ustilago ma
13	753	30.3	1058	10 Q9SJU0	Q9sj00 arabidopsis
14	748.5	30.1	1033	10 Q9LHL9	Q9lhl9 arabidopsis
15	737.5	29.7	699	5 Q8MPT8	Q8mpt8 caenorhabdi
16	735.5	29.6	747	11 Q8BNH4	Q8bnh4 mus musculus

17	735.5	29.6	757	11 Q8OU27	Q8ou27 mus musculu
18	733.5	29.5	443	11 Q8CGJ1	Q8cgj1 mus musculu
19	729.5	29.4	890	10 Q8W5R6	Q8w5r6 arabidopsis
20	728	29.3	701	11 Q7TSZ7	Q7tsz7 mus musculu
21	727.5	29.3	956	10 Q9C7B9	Q9c7b9 arabidopsis
22	727	29.3	997	10 Q9FG03	Q9fg03 arabidopsis
23	725	29.2	782	5 Q19633	Q19633 caenorhabdi
24	725	29.2	782	5 Q9GV92	Q9gv92 caenorhabdi
25	724	29.2	959	10 Q8S950	Q8s950 nicotiana t
26	720.5	29.0	735	5 Q9U0D5	Q9u0d5 tetrahymena
27	720	29.0	408	11 Q8OUK1	Q8ouk1 mus musculu
28	717.5	28.9	954	10 Q9AMW8	Q9amw8 oryza sativ
29	717.5	28.9	1193	10 Q7X7H8	Q7x7h8 oryza sativ
30	716.5	28.9	408	4 Q8LWH8	Q8lwh8 homo sapien
31	716.5	28.9	677	5 Q9VRK9	Q9vrk9 drosophila
32	716.5	28.9	699	13 Q98T11	Q98t11 xenopus lae
33	714	28.8	836	5 Q7YUC7	Q7yuc7 tetrahymena
34	713.5	28.7	1127	4 Q86TN3	Q86tn3 homo sapien
35	712.5	28.7	671	4 Q86XX7	Q86xx7 homo sapien
36	710.5	28.6	819	11 Q7TQ66	Q7tq66 mus musculu
37	710.5	28.6	974	10 Q8S905	Q8s905 arabidopsis
38	710.5	28.6	1231	11 Q80YP3	Q80yp3 mus musculu
39	707	28.5	160	11 Q35059	Q35059 mus musculu
40	706.5	28.5	744	13 Q93478	Q93478 xenopus lae
41	705.5	28.4	1037	10 Q3FH58	Q3fh58 arabidopsis
42	705	28.4	1121	10 Q8SV18	Q8sv18 arabidopsis
43	702.5	28.3	671	5 Q8MPT7	Q8mpt7 caenorhabdi
44	702.5	28.3	672	5 Q9GV93	Q9gv93 caenorhabdi
45	700	28.2	909	10 Q9XI03	Q9xi03 arabidopsis

ALIGNMENTS

RESULT 1

ID	Q7TPX4	PRELIMINARY;	PRT;	549 AA.
AC	Q7TPX4;			
DT	01-OCT-2003 (TREMBlrel. 25, Created)			
DT	01-OCT-2003 (TREMBlrel. 25, Last sequence update)			
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)			
DE	Hypothetical protein (Fragment).			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Egg;			
RX	MEDLINE=22389257; PubMed=12477932;			
RA	Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J., Marra M.A.;			
RA	"Generation and initial analysis of more than 15,000 full-length human			
RA	and mouse cDNA sequences."			
RA	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
RA	[2]			
RA	SEQUENCE FROM N.A.			
RA	STRAIN=C57BL/6J; TISSUE=Egg;			

RA Strausberg R.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC052843; AAH52843.1; --
 KW Hypothetical protein.
 FT NON TER 549 549
 SQ SEQUENCE 549 AA; 62899 MW; B707C97DC9FB38D3 CRC64;

Query Match 81.4%; Score 2021.5; DB 11; Length 549;
 Best Local Similarity 80.3%; Pred. No. 1.7e-116; Indels 9; Gaps 2;
 Matches 396; Conservative 35; Mismatches 53;

QY 1 AEEGAVAVCVVRPLNSREELGETAQVYKTKDNNVIYQVDGSKSFNFRVFGNETTKN 60
 DB 2 AEEASVAVCVVRPLNSREELGEATHIYKTKDNLAIQSDGKSFQFDRVDSNETTKN 61
 QY 61 VYEEIAAPIIDSAIQYNGTIFAYGQTASGKTYTMGSEDLGVIPRAIHDIQKIKKFP 120
 DB 62 VYEEIAVPIISSAIQYNGTIFAYGQTASGKTYTMGSEDLGVIPRAIHDIQKIKKFP 121
 QY 121 DREFLRVSMYELNETITDLCGTOKMKPLIREDVNRNYYVADLTVEEVYTTSEMAKW 180
 DB 122 EREFLRVSMYELNETITDLCNAQKKKPLIREDNRYVYSDLTVEEVYTTSEMAKW 181
 QY 181 ITGKESRHYGETKMNQSRSSHTIFRMILESREKGEPSNCEGSKVSHLNLVDLAGSER 240
 DB 182 LATGKKNHYGITKMNQSRSSHTIFRMILESREKGEPSNCEGSKVSHLNLVDLAGSER 241
 QY 241 AAGTGAAGVRIKEGCNINRSILFGLGVIKLSDGVGGINYRDSKLTILQNSLGGNPK 300
 DB 242 AAGTGAAGVRIKEGCNINRSILFGLGVIKLSDGVGGINYRDSKLTILQNSLGGNPK 301
 QY 301 TRICTITPVSFDETLTALQASTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKQLEE 360
 DB 302 TRICTITPVSFDETLTALQASTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKQLEE 361
 QY 361 VSLTRAQAMEKDLQALKEEDKLLQVQNEKIENLRMLVTSSSLTLOQLKQLEER 420
 DB 362 VNTKTRAQAMEKDLQALKEEDKLLQVQNEKIENLRMLVTSSSLTLOQLKQLEER 421
 QY 421 TWLGKINKMKNVADQFNPTITTKTKLSINLLREI-----DESVCSESDVFSNT 474
 DB 422 TWCYG--KMDSDNYEKFVPTSIITRKRTSVTSLENSLMKFGESAASEFEMLNNT 478
 QY 475 LDTLSELEWNPAT 487
 DB 479 LESLAEEVWSSAT 491

RESULT 2
 ID O42263 PRELIMINARY; PRT; 2954 AA.
 AC O42263;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Kinesin-related protein.
 GN XCENP-E laevis (African clawed frog).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OC NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98028574; PubMed=9363944;
 RA Wood K.W., Sakowicz R., Goldstein L.S., Cleveland D.W.;
 RT "CENP-E is a plus end-directed kinetochore motor required for
 metaphase chromosome alignment."
 RL Cell 91:357-366(1997)
 DR EMBL; AF027728; AAC60300.1; --
 DR F01; T14156; T14156.
 DR HSSP; P33176; 1BG2.
 DR GO; GO:0005871; C:kinesin complex; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003774; F:Motor activity; IEA.
 DR GO; GO:0007017; F:Microtubule-based process; IEA.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF02225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS00657; KINESIN MOTOR DOMAIN2; 1.
 DR SEQUENCE 2954 AA; 339964 MW; 439804ED0E592679 CRC64;

Query Match 63.8%; Score 1584; DB 13; Length 2954;
 Best Local Similarity 64.0%; Pred. No. 1.4e-88;
 Matches 320; Conservative 68; Mismatches 96; Indels 16; Gaps 6;

QY 1 AEEGAVAVCVVRPLNSREELGETAQVYKTKDNNVIYQVDGSKSFNFRVFGNETTKN 60
 DB 2 SEGDAVKVCVRPLNQREQ--GDANLQWKAGNNTISQVDGTSKSFNFRVFNHSTSQ 59
 QY 61 VYEEIAAPIIDSAIQYNGTIFAYGQTASGKTYTMGSEDLGVIPRAIHDIQKIKKFP 120
 DB 60 IYQEIAPVPIRSALQYNGTIFAYGQTASGKTYTMGTPNSLGLIIPQAIQEVFKIQEIP 119
 QY 121 DREFLRVSMYELNETITDLCGTOKMKPLIREDVNRNYYVADLTVEEVYTTSEMAKW 180
 DB 120 NRREFLRVSMYELNETIVKLLCDDRKKKPLEIREDFNRNYYVADLTVEELVMVPEHVIQW 179
 QY 181 ITGKESRHYGETKMNQSRSSHTIFRMILESREKGEPSNCEGSKVSHLNLVDLAG 237
 DB 180 IKKGEKNHYGETKMDHSSRSHTIFRMIVESDRNDPTNSENCDGAVMVSHLNLVDLAG 239
 QY 238 SERAAGTGAAGVRIKEGCNINRSILFGLGVIKLSDGVGGINYRDSKLTILQNSLGG 297
 DB 240 SERASQTGAAGVRIKEGCNINRSILFGLGVIKLSDGVGGINYRDSKLTILQNSLGG 299
 QY 298 NPETRICTITPVSFDETLTALQASTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKQ 357
 DB 300 NAKTVIITPVSFDETLTALQASTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKQ 359
 QY 358 LE--EVSLETRAQAMEKDLQALKEEDKLLQVQNEKIENLRMLVTSSSLTLOQLKAK 415
 DB 360 LENLESSETKAQAKAKEHTQLIAETIKQLKEREDRIWHLTNIVWASSQES--QDDQVK 418
 QY 416 RKRVTWCLGINKMKNVADQFN-----IPTNLTTKTKLSINLLREIDSVCSSESVF 471
 DB 419 RKRVTWAPGKIQNSLHSGVSDFDMLSRGPNFSKKAKFSDMPSPFDDSDVCTEFSDF 478
 QY 472 SNTLDTLS----ETWNPAT 487
 DB 479 DDALSMDSDNGIDAEWNLAS 498

RESULT 3
 ID Q94HV9 PRELIMINARY; PRT; 807 AA.
 AC Q94HV9;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Kinesin motor protein (kin2), putative.
 GN T4M14.11.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Town C.D., Haas B.J., Wu D., Maiti R., Hannick L.I., Chan A.P.,
 RA Tallon L.J., Rooney T., Utterback T.R., VanAken S.E., Feldblum T.V.,
 RA White O., Fraser C.M.;
 RA "Arabidopsis thaliana chromosome 1 BAC T4M14 genomic sequence.";

[illegible][illegible]

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Putative kinesin-like centromere protein.
GN F14P13.22
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Roming C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RA "Arabidopsis thaliana chromosome III BAC F14P13 genomic sequence.";
RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AC009400; AAF02823.1; -;
DR HSP; P33176; 18G2.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; Kinesin; 1.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
SQ SEQUENCE 459 AA; 51872 MW; 768487E9B7038E6A CRC64;
Query Match 34.5%; Score 856; DB 10; Length 459;
Best Local Similarity 50.7%; Pred. No. 8.8e-45;
Matches 194; Conservative 62; Mismatches 97; Indels 30; Gaps 8;
Qy 49 DRVPHGNETKNVVEETAAPIIDSAIQYNGTIFAYGCTASGKTYTMMGSDHLGVIPRA 108
Db 45 DRIFREDCKTVQVYEARTKEIVSAARVGFNGTVFAYGQTNSGKTHMRGSPIEGVIPLA 104
Qy 109 IHDIFQKIKPPDFREFLLRYSMEIYNETITDLCCGTQKMKPLIREDVNRNVYADLTE 168
Db 105 VHDLFDTIYQDASREFLLRYSMEIYNEEDINDLI--APEHRKLOIHENLEKGI FVAGLRE 162
Qy 169 EVVYSEMALKWITKGEKSRHYGTHKQNRSSRSHITFRMILEGREK----GEPSNCEGS 224
Db 163 EIVASPPQVLEMEFGEHSHRIGETNNMNLYSRSHITFRMILESRQKMQDEGVNSCD-A 221
Qy 225 VKVSHNLVDLAGSRAAQTCAAGVRLKGCNINRSLFILGOVTKLSDG--QVGGFINY 282
Db 222 VRVSVNLVDLAGSRAAQTCAAGVRLKGCNINRSLFILGOVTKLSDG--QVGGFINY 281
Qy 283 RDSKLTILQNSLGNPKTRICIT--PVSFDETLTALQFASAKYMKNTPYVNEVSTD 340
Db 282 RDSKLTILQALGNANTAIICNTLAPIHADETKSLQFASRALRVNCAHVNEILTD 341
Qy 341 EALKRYRKEIMDLKKQL-----EVSLIETR-----AQAMEKDLAQLEEKDLIQ-- 386
Db 342 AALLKQKQKETEELRSKLTSHSDHSEBEILNRLNTLKSLELERIALALEEKEKAQAQ 401
Qy 387 -----KVQNEKIENLTMVTS 404
Db 402 RERVLOEQAKKIKNLSMVLJLN 424
RESULT 6
ID Q9LQ62 PRELIMINARY; PRT; 888 AA.
AC Q9LQ62;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE T30E16.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Roming C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RA "Arabidopsis thaliana chromosome III BAC F14P13 genomic sequence.";
RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AC009400; AAF02823.1; -;
DR HSP; P33176; 18G2.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; Kinesin; 1.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
SQ SEQUENCE 459 AA; 51872 MW; 768487E9B7038E6A CRC64;
Query Match 34.0%; Score 844.5; DB 10; Length 888;
Best Local Similarity 40.0%; Pred. No. 1.1e-43;
Matches 220; Conservative 70; Mismatches 169; Indels 91; Gaps 13;
Qy 6 VAVCVVRPLNSREESIGETAQVYVKYDNNVI-----Y 38
Db 4 ICVAVVRP-----PAPENGASLWKVEDNRISLHSLDTPITTASHAFVSGISISTDLI 57
Qy 39 QVDSKSKFNF-----DRVPHGNETKNVVEETAAPIIDSAIQYNGTIFAYGCTA 88
Db 58 EIVSLFLFSGVGVYFLLFDHDFDESSNVSVELTQDIHAHAEVGFNTAFAYGQIS 117
Qy 89 SGKTYTMMGSDHLGVIPRAIHDFQKIKPPDFREFLLRYSMEIYNETITDLCCGTQK 148
Db 118 SKGTYTMMGSDHLGVIPRAIHDFQKIKPPDFREFLLRYSMEIYNETITDLCCGTQK 147
Qy 149 KPLIREDVNRNVYADLTEEVYSEMALKWITKGEKSRHYGTHKQNRSSRSHITFRM 208
Db 176 QRLQTHHELRGVFVAGLKEEIVDSAEQILKLDSDGEVNRHFGTNNMNVHSRSHITFRM 235
Qy 209 -----ILESREKGEPSNCEGSVKVSHNLVDLAGSRAAQTCAAGVRLKGCN 256
Db 236 VRFSYERDLLLLVIESR--GKDNSSDAIRVSVNLVDLAGSERIAKTGAGVRLQEGKY 293
Qy 257 INRSILFILGOVTKLSDG--QVGGFINYRDSKLTILQNSLGNPKTRICITTPVS--FD 313
Db 294 INKSMILGNVINKLSDSTKLRAHPIPYRDSKLTILQALGNKANTCIITIAPEEHIE 353

```

QY 314 ETLTALQFASAKYKNTPYNEVSTDEALLKRYRKEIMDLKQLLEVSLETRAQ----- 368
Db 354 ESKTLOFASRAKRITCAQVNEILTTAALLKROKLEIEELURMKLQGSBAVLQOEILNL 413
QY 369 -----AMEKDQLAQLLEEDLLQKVQNE-----KIENLTRMLVTSSSLTLQOELKA 414
Db 414 SNQMLKVELECEERLKTOLLEBKRKQKQENCICKEQQMKIENLNN-FVNSDF----- 464
QY 415 KRKRVTWCLGKINKMKNNSVADQFNPTNTTKTKLKSINLLREIDSVCSGSDVFSNT 474
Db 465 KRQNSDFIIRKPTPDGLCNVNDTSDVBTGTCFKSASRFSVVARSNYSGLSDFSPMVHS 524
QY 475 L-DTLSELEW 483
Db 525 LGDVAEDTW 534

RESULT 7
ID Q869B8 PRELIMINARY; PRT; 1885 AA.
AC Q869B8;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Kinesin-related protein K4.
GN K4.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
EN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98359834; PubMed=9693369;
RA de Hostos E.L., McCaffrey G., Sugang R., Pierce D.W., Vale R.D.;
RT "A developmentally regulated kinesin-related motor proteins from
RT Dictyostelium discoideum."
RL Mol. Biol. Cell 9:2093-2106(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Suyama E., Sutoh K.;
RT "Kinesin-related proteins from Dictyostelium."
RT Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB102780; BAC56912.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005875; C:microtubule associated complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR001752; F:motor activity; IEA.
DR InterPro; IPR003345; M_repeat.
DR Pfam; PF00225; kinesin; 1.
DR Pfam; PF02370; M; 6.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
SQ SEQUENCE 1885 AA; 218216 MW; 15686AEDB3B007EC7 CRC64;

Query Match 32.3%; Score 801; DB 5; Length 1885;
Best Local Similarity 39.7%; Pred. No. 1.4e-40;
Matches 193; Conservative 79; Mismatches 144; Indels 70; Gaps 9;

QY 2 EBGAVACVRVPLNSRESISGETAQVYKWTNNVI-YQVDSKSFNEDRVFGHNETKN 60
Db 19 ELNKIKVAIRVPLNSRE--LGIDQKIPWSISKDTLSLQNFNINFTYDVFQIDSNITD 76
QY 61 VVEETAAPIDGALQNGYCTIFAYGQTASGKYTWGMSDHLGVIPRAIHDFOKIK-KF 119
Db 77 VYNAKLSIVNSLNGINGTIFAYGQTSKGFSGMGTSGIPGILKLSIKDIFKSIEDSI 136
QY 120 PDRPEFLRVSYMEIYNETITDLGGT-QKQKPLIINEDVRNRYVADLTVEVYVYSEMAL 178
Db 137 LEKDYLKVSYLEIYNEEIKOLLNPTISNKKLIKIHEDIYKGVGVANLKEEIVISPDQTF 196

```

```

QY 179 KWIITGKSRHYGTEKMNQSRSSHTTFRMILESRKEGPEPCNCEGSKVXVSHNLVLDLAGS 238
Db 197 ALMNFGEERRHIGSTWMDSSSRSHITFRMQIOSTCKQ-----NGTIQMSLTIVLDLAGS 251
QY 239 ERAAQTGAAGVRLKEGNCINRSFILQVYIKLSDGQGVGFNYRDSKLTILQNSLGGN 298
Db 252 ERVSTGAEGVRLKEGTHINKSLMTLSKLSKSEKTTQQHPYRDSKLTILQPSLGGN 311
QY 299 PKRIICITIPVSF--DETLTALQFASAKYKNTPYNEVSTDEALLKRYRKEIMDLK 356
Db 312 SKTAILCTITPAITHQEEISTLQFAKRAKVKTKYKINQINQVADANTMLKKYSEILELQN 371
QY 357 QLEEVSLTRAQAMEKDXLAQLLEEKDQLQVQNEKIENLTRMLVTSSSL----- 406
Db 372 QL-----VKSEINSLLRNTISTQEISSNNFKLGWK 402
QY 407 -----TLQQLKAKRKRRVTW-----CLGKINKMKNNSVADQFNPTNTT 447
Db 403 RFNDATIGGSLINENKKKKRNTLDPFSLYLLKDKIIRKRSNKQIKKIKNSENNIS 462
QY 448 KTHKLS 453
Db 463 SSSNS 468

RESULT 8
QYVKIO PRELIMINARY; PRT; 2013 AA.
ID QYVKIO
AC QYVKIO;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Cmet protein
GN CMET OR C6392.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

```

[illegible]

OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;	
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=20351410; PubMed=10893249;	
RA	Yucel J.K., Marzalek J.D., McIntosh J.R., Goldstein L.S.B.,	
RA	Cleveland D.W., Philp A.V.;	
RT	"CENP-meta, an Essential Kinetochores Kinesin Required for the	
RT	Maintenance of Metaphase Chromosome Alignment in Drosophila.";	
RL	J. Cell Biol. 150:4-12(2000).	
DR	EMBL; AF220353; AAF32355.1; -.	
DR	HSSP; P33176; 18G2.	
DR	FLYBASE; FBGN0040232; cmet.	
DR	GO; GO:0005699; C:kinetochores; IDA.	
DR	GO; GO:0007080; P:mitotic metaphase plate congression; IMP.	
DR	InterPro; IPR001752; kinesin_motor.	
DR	Pfam; PF00225; kinesin; 1.	
DR	PRINTS; PR00380; KINESINHEAVY.	
DR	SMART; SM00129; KISC; 1.	
DR	PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.	
DR	PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.	
SQ	SEQUENCE 2244 AA; 257992 MW; _FA6AA3B2A541ADE0 CRC64;	

Query Match	31.4%;	Score 780.5;	DB 5;	Length 2244;
Best Local Similarity	40.8%;	Pred. No. 3.2e-39;		
Matches 203;	Conservative 76;	Mismatches 148;	Indels 71;	Gaps 16

Qy	5	AVAVCVRRVRLNRSREELGETAQYVWKT-DNNVYQVDG-SKSFNDFRVFHGNETTKNY	62
Db	8	SIQVICVRPCPEGLTSL-----WQVKEGRSIHLADSHAEPYVDFVDFDEGASNQEVF	60
Qy	63	ETAAPIDSAIQYNGTIFPAYGTASGKTYTMGSEDLHGVIPRAIHDFOKIKKFPDR	122
Db	61	DRMARHIVHACMQGFGNTIFPAYGTSKGKTYTMGDEQNPQVWVLAAKEIFQOISSETER	120
Qy	123	EFLRLVSYMEIYNTFTIDLCGTCQMKPLIIRVDNRNVVYADLTVEVVYVTSSEM-ALKWI	181
Db	121	DFLLRVGYEILYNEKIYDLL--NKKQDLKHESNGIYVNV--CECCIITSEVDLLRL	176
Qy	182	TKGEKSHYGETKMQRSSRSHITFRMLSEKGEPSNCGSVKSHNLNVDLAGSERA	241
Db	177	CLGNKERTVGETNNERSRSHAFKIIIESR-KSDHSD-DDAVIQSVNLNVDLAGSERA	234
Qy	242	AQTGAACVRKLEGCNINRSFILGVTKLSDGQVGGINVYRDSKLTIRLQNSLGNPKT	301
Db	235	DQIGARGARLKBGGHINKSLFLSNVLSSENADRNFTNRDSKLTIRLQNSLGNAPT	294
Qy	302	RIICITTPVSFDETLTALQFASTAKYMKNTPVYNEVSTDEALLKRYRKEIIMDLKQLEEV	361
Db	295	SIICIKPSIMBESQSTLSPATRAKKIRIKPQVNEWVS DATMVKRLEREIKVLKDKLAE-	353
Qy	362	SLETRAQAEKDQLAQLLEKDLQKVNEXIENLTM-----LVTSSTLTQOELKA	414
Db	354	-----EER-----KNEQQKVEHLERQIKHDMHKIIICGHSLS-----DKG	388
Qy	415	KRKRVRVTC-----LGKINKMKNVADQF-----NIPTNITTKTHKL-----SI	454
Db	389	QOKRRRWCTFASGSHLEAETGTEDRIDQFPKVSHLPKVPFFHTSNAGKWDNI PKTI	448
Qy	455	NLLREIDESVCSSESDVFS	472
Db	449	NILGSLD--IGTESNSIS	464

RESULT 10	
ID Q8RW4	PRELIMINARY; PRT; 1055 AA.
AC Q8RW4	
DT 01-JUN-2002 (Tremblrel. 21, Created)	
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)	


```

QY 301 TRIICITTPV--SFDETTLTALQFASAKYMKNTPTVYNEVSTDEALLKRYRKEIMDLKKQL 358
Db 386 VSLICITTPASSSEETHNLTKASRAKSIIEIVASRNQIIDKSLIKKYQREISTKLKL 445
QY 359 EVV-----SLETRAQAMEKQ--LAQLLEFKDLQKQVNEKIEINLRLMLVTSS 405
Db 446 DQLRRGLMVGVSHEELMSLQKQLEGGQVQKQSRLEEEBAKAALMSRIQKLTKLIVST- 504
QY 406 LTLQOELKAKRKRVTWCLGKINKMKNSYADQFNITNITTKTKLSINLLREID-ESV 464
Db 505 -----KNSIPGSGDIP-----THQRSAGKDDKFDLSL 533
QY 465 CSBSDFV---SNTLDTLSE 480
Db 534 LLESNDNLGSPSSTLALLSE 552

RESULT 12
ID P87198 PRELIMINARY; PRT; 1459 AA.
AC P87198;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Kinesin motor protein.
GN KIN1.
OS Ustilago maydis (Smut fungus).
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_TaxID=5270;
[1]
SEQUENCE FROM N.A.
MEDLINE=97361828; PubMed=9218789;
EX Lehmler C., Steinberg G., Snetseelaar K.M., Schliwa M., Kahmann R.,
RA Bolker M.;
RA "Identification of a motor protein required for filamentous growth in
RT Ustilago maydis.";
RL EMBO J. 16:3464-3473 (1997).
DR EMBL; U92844; AAB63336.1; -.
DR PIR; T30196; T30196.
DR HSP; P17119; 3KAR.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; Kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN_MOTOR_DOMAIN2; 1.
SQ SEQUENCE 1459 AA; 156932 MW; 281374B0100F5CD0 CRC64;

Query Match 30.6%; Score 759; DB 3; Length 1459;
Best Local Similarity 38.2%; Pred. No. 3.8e-38;
Matches 191; Conservative 71; Mismatches 128; Indels 110; Gaps 14;

QY 6 VAVCVVRPLNSREESLGETAQVYVKTDN--NVLYQVD----- 41
Db 237 VVVCVRMRP--SRASSDSSEASV--WNCSEKNRIFPTHEHPAIAKRTTSSERAGAGASIA 293
QY 42 -----GSKSFNFDRVFHGNETHKKNVYEEIAAIPDISAIGYNGTIFAYGQ 86
Db 294 AAPSSHDLHDHEDPTSSYVHFQDKLITGAQITDDMYHSHIAPVVEAAVEGYNGTVFAYGQ 353
QY 87 TASGKTYNMGSEDLHGVIPRAIHDIPOKIKKFPDRFLLRVSYMEINETHLLCTGQ 146
Db 354 TGSCKTHTMSGDAEPGVPRAVEQIFQMIKDEPDRELLRVSYLEIYNETLKKLLA--- 410
QY 147 KMKPL-----IREDVNRNVVADLTSEVW 171
Db 411 PLPLITGTGSSLQTTDRPASPIKGGSSHAAGQSQCILRIEDQKSRVITGLREIV 470

```

```

QY 172 YTSMALKWITTKGEKSRHVGGETKRONRSSRSHTIFRMILESREKGPSCGSKVSHLN 231
Db 471 TDANTVLCILQQRDERHVGATDWNERSRSRSHCVFQLTIESRPAFSAKSE--VRISQLN 528
QY 232 LVTLASERAAQTGAAGVRLKEGCNTRSLFILGOVTKKLSDGOVG3--FINYRDSKLTR 289
Db 529 LIDLASERAA---SQAERKEGAFINKSLTLTGTVIKUTEPEVNGDAHIPIRDSKLTR 585
QY 290 ILQNSLGGNPKTRIICTITPVS--FDETTLTALQFASAKYMKNTPTVYNEVSTDEALLKRY 347
Db 586 ILTSLSGNARIIVICTLSPDTEHANETLSLTFGKCKLVVITAKGTAMDDKALLQKY 645
QY 348 RKEIMDLKKOLEE-----VSLTRAQAMEK-DQLAQLEB-----EKDLQKQV 389
Db 646 RKELDALRAKLEANGPSPNEVMTIVSABASKESQQLDQLAQOKEAAQREVEDMQKR 705
QY 390 NE---KIENTRLMLVTSSSL 406
Db 706 SHLKAQIEHTRILTLTSQSV 725

RESULT 13
QY Q9SJUU PRELIMINARY; PRT; 1058 AA.
AC Q9SJUU; Q94BQ1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative kinesin heavy chain.
GN AR2G21380.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA Barnstead M.E., Mason T.M., Bowman C.L., Renning C.M., Benito M.-I.,
RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,
RA Fraser C.M., Venter J.C.;
RA Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Dale J.M., Gibson H.A., Goldsmith A.D., Jiang P.X., Lee J.M.,
RA Quach H.L., Tang C., Toriumi M., Yu G., Bowser L., Carninci P.,
RA Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RA "Full length cDNA of gene F3K23.14/At2G21380 (GI:4567271).";
RA Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
RA Quach H.L., Tang C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,
RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
RA Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RA "Arabidopsis Open Reading Frame (ORF) Clones.";
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AC006841; AAD23684.2; -.
DR EMBL; AY039966; AAK64143.1; -.
DR PIR; E84600; E84600.
DR HSP; P17119; 3KAR.
DR GO; GO:0005871; C:kinesin complex; IEA.

```


[1]
RRP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RC MEDLINE=99069613; PubMed-9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RP STRAIN=Bristol N2;
RA Nelson J., Wohldmann P.;
RA "The sequence of C. elegans cosmid M02B7";
RT Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RL [3]
RN SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RP STRAIN=Bristol N2;
RA Waterston R.;
RA Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U70851; AAM97997.1...
DR WormRep; M02B7.3b; CE3158.
DR GO; GO:0003871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; Kinesin; I.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; Kisc; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS50067; KINESIN MOTOR DOMAIN2; 1.
DR SEQUENCE 699 AA; 78779 MW; 8A774E3EF3A07813 CRC64;
SQ

Query Match 29.7%; Score 737.5; DB 5; Length 699;
Best Local Similarity 38.5%; Pred. No. 3.1e-37;
Matches 200; Conservative 174; Indels 67; Gaps 17;

Qy	5	AVAVCVVRPLNSREESLGETAQVYWKTDNNVIYOV-----DG-SKSFNEDRVFHNETT	58
Dd	4	SVRVAVRCPFNQEKDNLTTLCV-GMTPN--VGQVNLPADGAAKOTFGAYFMDSTG	60
Qy	59	KNVVEETAAPIDSAIOGYNGTIFAYGQTASGKTYYTMGSBD---HLGVIPRAIHDPK	115
Dd	61	EQIVNDIVFPVENVIEGYNGTVFAYGQTSGKTFSMOGIETIPAQRGVIIPRAFPHIFTA	120
Qy	116	IKKPPDREFLLRSVMRYNETIIDLLCGTKMKPLIIREVDNRNVYADVTEEVYTSE	175
Dd	121	TATTENVKFLVHCYSLEYNEVEDLLIGADNKQK-LRIEQPDGRGVYAGLSMHVCHDVP	179
Qy	176	MALKWITKGSRHYGHGTKNQRSRSHTIIRFMILESRKEGPSCEGSVKVSHNLVDL	235
Dd	180	ACKELMTRGFNHRHVGAATLMNKDSRSHSIFTVYEGMTE-----TGISRMGNKLNLVDL	233
Qy	236	AGSERAAQTGAAGVRLKGGCNINSLFLTGQVKKLSDGQVGGFNYNRDSKLTRILQNSL	295
Dd	234	AGSERQSKTGATGRLXKATKINLSALGNVISALVDGK-SKHPIPYRDSKLTRLQLDSL	292
Qy	296	GGNPKTRIICHTPVVS--FDFTLALQPASTAKYMNKPTPNYNEVSTDALLKRYRKEMD	353
Dd	293	GGNTKTIIMACVSPSSNDYDETLSTLRVYANAKNIKKNKPTINEDPKD-ALLREYOEEIAR	351
Qy	354	LKKOLEVSLSTRQAAM-----EKDLAQALLEE--KDILLQKVQNEKIENLTRMLVTSS	404
Dd	352	LKSVMVPGAVGVGAPADAFAIERERKKLRREEFEAMNDLRGEYREQ-----TS	401
Qy	405	SLTLQQELLKAKRK--RRVTWCGLGIN----RWKSNYNADQF-----NIP	442
Dd	402	KAEIQKOLLESRADYERANANDLNLPPEAAKKIQLQLODFIGGEEAGNTQLKQKRMKQL	461
Qy	443	TNIITTKTKISINLLREIDESVCSESDFNSVNTLDTLSEI	481
Dd	462	KEAETKTOKLAALNVHDKDEL---QVYSTTQEKLDAV	497

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:26:01 ; Search time 20.4051 Seconds

(without alignments)
4694.096 Million cell updates/sec

Title: US-10-045-631b-88_COPY_2_340

Perfect score: 1744

Sequence: 1 AEEGAVAVCVVRPLNSREE.....QFASAKYMKNTFYNEVNST 339

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_23Jan04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1744	100.0	2633	4	ABG06505	Abg06505 Novel hum
2	1744	100.0	2663	4	AAm39097	AAm39097 Human pol
3	1635.5	93.8	2688	4	AAm40883	AAm40883 Human pol
4	1299.5	74.5	2954	2	AAy01632	AAy01632 Amino aci
5	781	44.8	348	4	ABU53208	ABU53208 Human cel
6	781	44.8	366	4	ABU53125	ABU53125 Intracell
7	707	40.5	2013	4	ABG62322	ABG62322 Drosophil
8	668	38.3	473	5	ABG70992	ABG70992 Human tar
9	668	38.3	522	5	ABG70991	ABG70991 Human HsK
10	668	38.3	1232	7	ADD49338	ADD49338 Human lun
11	658	38.3	1232	7	ADD18924	ADD18924 Human dis
12	658	37.7	1232	7	ADD49937	ADD49937 Human kin
13	658	37.7	1232	7	ADD49932	ADD49932 Human lun
14	652	37.4	1232	5	ABG70990	ABG70990 Human HsK
15	651	37.3	1234	5	ABG70993	ABG70993 Human HsK
16	649.5	37.2	1029	5	AAE17786	AAE17786 Human kin
17	649.5	37.2	1029	7	ADC10190	ADC10190 Human MOV
18	646.5	37.1	677	4	ABB65183	ABB65183 Drosophil
19	641.5	36.8	1038	5	AAm48337	AAm48337 Human kin
20	621.5	35.6	1034	3	AAQ31112	AAQ31112 Arabidops
21	621.5	35.6	1069	3	AAQ31111	AAQ31111 Arabidops
22	621.5	35.6	1121	3	AAQ31110	AAQ31110 Arabidops
23	620.5	35.6	329	5	AAE17787	AAE17787 Human kin
24	618	35.4	784	2	AAy06618	AAy06618 Thermomyc
25	615.5	35.3	411	2	AAW72745	AAW72745 Drosophil

ALIGNMENTS

RESULT 1

ABG06505

ID ABG06505 standard; protein; 2633 AA.

XX AC

XX ABG06505;

XX AC

XX DT 13-FEB-2002 (first entry)

XX XX

DE Novel human diagnostic protein #6496.

XX Human;

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX XX

OS Homo sapiens.

XX XX

PN WO200175067-A2.

XX XX

XX 11-OCT-2001.

XX XX

PF 30-MAR-2001; 2001WO-US008631.

XX XX

PR 31-MAR-2000; 2000US-00540217.

XX XX

PR 23-AUG-2000; 2000US-00649167.

XX XX

XX (HYSE-) HYSEQ INC.

XX PA

XX PI Dmanac RT, Liu C, Tang YT;

XX XX

XX WPI; 2001-639362/73.

XX DR

XX N-PSDB; AAs70692.

XX PT

XX New isolated polynucleotide and encoded polypeptides, useful in

XX PT

XX diagnostics, forensics, gene mapping, identification of mutations

XX PT

XX responsible for genetic disorders or other traits and to assess

XX PT

XX Claim 20; SEQ ID NO 36864; 103pp; English.

XX PS

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)

XX CC

XX sequences. (I) is useful as hybridisation probes, polymerase chain

XX CC

XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

XX CC

XX and in recombinant production of (II). The polynucleotides are also used

XX CC

XX in diagnostics as expressed sequence tags for identifying expressed

XX CC

XX genes. (I) is useful in gene therapy techniques to restore normal

XX CC

XX activity of (II) or to treat disease states involving (II). (II) is

XX CC

XX useful for generating antibodies against it, detecting or quantitating a

XX CC

XX polypeptide in tissue, as molecular weight markers and as a food

XX CC

XX supplement. (II) and its binding partners are useful in medical imaging

Human polypeptide SEQ ID NO 5814.

Human; neotrophic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia.

OS Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.

XX 23-DEC-1999; 99US-00471275.

XX 21-JAN-2000; 2000US-00488725.

XX 25-APR-2000; 2000US-00552317.

XX 20-JUN-2000; 2000US-00598042.

XX 19-JUL-2000; 2000US-00620312.

XX 03-AUG-2000; 2000US-00653450.

XX 14-SEP-2000; 2000US-00662191.

XX 19-OCT-2000; 2000US-00693036.

XX 29-NOV-2000; 2000US-00727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;

XX Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

XX N-PSDB; AAI60039.

XX Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries.

XX Example 2; SEQ ID NO 5814; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the encoded polypeptides (AAM38642-AAM42213) with neotrophic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders. Note: The sequence data for this patent did not form part of the printed specification

XX Sequence 2688 AA;

Query Match 93.8%; Score 1635.5; DB 4; Length 2688;

Best Local Similarity 95.3%; Pred. No. 1.1e-150;

Matches 326; Conservative 2; Mismatches 11; Indels 3; Gaps 3;

QY 1 AEEGAVACVVRPLNSRESLGETAQYWKTDNRVIVQVDSKSFNDRVFGHNETPKN 60

DB 23 AEEGAVACVVRPLNSRESLGETAQYWKTDNRVIVQVDSKSFNDRVFGHNETPKN 82

QY 61 VYEETAAPIIDSAIGYNGTTFAYGQTASGKTYTMMGSEDLGVTPRA-IHDIF-QKIKK 118

DB 83 VYEETAAPIIDSAIGYNGTTFAYGQTASGKTYTMMGSEDLGVTPRA-IHDIF-QKIKK 142

QY 119 -FPDRFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVYADLTVEEVYTSEMA 177

Db 143 VFDRFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVYADLTVEEVYTSEMA 202

QY 178 LKMITGKSRHYGETKMQRSRSHITFRMILESREKGFPSNCEGSKVSHLNLVDLAG 237

Db 203 LKMITGKSRHYGETKMQRSRSHITFRMILESREKGFPSNCEGSKVSHLNLVDLAG 262

QY 238 SERAATGAGVRLKEGCMNINSLFILGQVVKLSGQGVGFNYRDSKUTRILQNSLGG 297

Db 263 SERAATGAGVRLKEGCMNINSLFILGQVVKLSGQGVGFNYRDSKUTRILQNSLGG 322

QY 298 NPKTRIICITTPVSFDETLTALQFASTAKYMKNTPIVNEVST 339

Db 323 NPKTRIICITTPVSFDETLTALQFASTAKYMKNTPIVNEVST 364

RESULT 4

AAI01632

ID AAY01632 standard; protein; 2954 AA.

AC AAY01632;

DT 22-JUN-1999 (first entry)

XX Amino acid sequence of centromere-associated protein-E (CENP-E).

XX CENP-E; centromere-associated protein-E; ATPase activity;

XX plus end-directed microtubule motor activity; chromosome congression;

XX microtubule binding activity; chromosome movement; mitosis;

XX cell proliferation; tumor; metastasis; vascular malfunction;

XX inflammatory disease; immune disease; angiogenesis; hypertension;

XX restenosis; fungal infection; selective herbicide; fungicide;

XX insecticide; plant growth regulator; activator; cancer cell marker.

XX Xenopus sp.

XX WO9913061-A1.

XX 18-MAR-1999.

XX 10-SEP-1998; 98WO-US019231.

XX 11-SEP-1997; 97US-0058645P.

XX (REGC) UNIV CALIFORNIA.

XX Wood KW, Sakowicz R, Goldstein LSB, Cleveland DW;

XX WPI: 1999-229233/19.

XX N-PSDB; AAX26819.

XX Centromere-associated protein-E and related nucleic acid.

XX Claim 5; Page 66-67; 77pp; English.

XX The present sequence represents CENP-E (centromere-associated protein-E) of Xenopus. The protein has at least one of plus end-directed microtubule motor activity, ATPase (adenosine triphosphatase) activity and microtubule binding activity. CENP-E is the motor that powers chromosome movement toward microtubule plus ends and is essential for congression of chromosomes during mitosis. Modulators of CENP-E can thus control cell proliferation. Agents that modulate CENP-E activity are lead therapeutic, bioagricultural and diagnostic agents, e.g. for treatment of unwanted cell proliferation (typical of many examples are tumors and metastases; vascular malfunction; inflammatory and immune diseases; angiogenesis; hypertension; restenosis; and fungal infections), also as plant-protection agents (selective herbicides, fungicides and insecticides) and plant growth regulators or activators for improving yields. CENP-E is also a diagnostic marker for dividing cells, including cancer cells

XX Sequence 2954 AA;

Query Match 74.5%; Score 1299.5; DB 2; Length 2954;

Best Local Similarity	73.5%;	Pred. No.	1.5e-117;		Gaps	2;	
Matches	250;	Conservative	42;	Mismatches	43;	Indels	5;
QY	1	AEGGAVACVVRPLNSRRESLGETAQVWKTIDNNVIYQVDGKSFNDRVFHGNETXN	60				
		: :					
Db	2	SEGDAVKVCVRVRLPIREQ--GDQANLQWKAGNTISQVDTGTSFNDFRVFNSHESQ	59				
		: :					
QY	61	VYEIEAIPITDSALIQVNGCHIFAYGTASGTYTWMGSEDLGVLPRAIHDIFOKIKKFP	120				
		: :					
Db	60	IYOEIIAPVILRSALQGYNGTFAYGTSSGKYITMGTTPMSLGITPAIOBGFVKLIQEIP	119				
		: :					
QY	121	DREFLLRVSWEIYNETITOLLCTQKKMLPIREDVNRNVYVADLTVEEVYTSEMALKW	180				
		: :					
Db	120	NREFLLRVSWEIYNETIVKLLCDDRKKPLEREDFRNVYVADLTVEELVMVPEHVIO	179				
		: :~::~::					
QY	181	ITKGKSRHYGETIQONQRSSHTIFRMILESRKEGPS---NCEGSVKVSHNLVDLAG	237				
		: :					
Db	180	IKKGKNRYHGETKWNDSRSSHTIFRMIVESRDNDPTNSECDGAVWSHLNLVDLAG	239				
		: :					
QY	238	SERAQAQTGAAGVRLKEGCINIRSLFILGOVTKLSDGQVGGINVRDSKLTRILQNSLGG	297				
		: :					
Db	240	SERASQTGAEGVRLKEGCINIRSLFILGOVTKLSDGQAQGFINVRDSKLTRILQNSLGG	299				
		: :					
QY	298	NPKTRICTITPVSFDETLTALQFASTAKYMKNTPPYNEV	337				
		: :					
Db	300	NAKTVICTITPVSFDETLSTLOFASTAKHYRNTPHYNEV	339				
		: :					

RESULT 5	
ABUS3208	
ID	ABU53208 standard; protein; 348 AA.
XX	
XX	
XX	ABU53208;
XX	
XX	
DT	14-APR-2003 (first entry)
XX	
DE	Human cell cycle-associated DXFzptes3_35b4 homologue #7.
XX	
XX	
KW	Human; gene therapy; vaccine; disease treatment; detection.
XX	
XX	
OS	Homo sapiens.
XX	
PN	WO200112659-A2.
XX	
XX	
PD	22-FEB-2001.
XX	
PF	18-AUG-2000; 2000WO-IB001496.
XX	
XX	
PR	18-AUG-1999; 99US-0149499P.
PR	28-SEP-1999; 99US-0156503P.
XX	
XX	
PA	(GEHU-) GERMAN HUMAN GENOME PROJECT.
XX	
XX	
PI	Wiemann S;
XX	
XX	
DR	WPI: 2001-327840/34.

```

SQ      Sequence 348 AA;
Query Match      44.8%; Score 781; DB 4; Length 348;
Best Local Similarity 51.1%; Pred. No. 6.6e-68;
Matches 178; Conservative 38; Mismatches 106; Indels 26; Gaps
17;

Qy      11 RVPLNRSRESLGETAQVYK-----KTDNNVYQVDGSKSFNDRFVHGNETKNVYEE 64
      ||||| : : : : : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db      1 RCPLNREINDCGSCVQVPPWGYKTVENGHGDSPHKSFTHDVFWNCTQEDVYDT 60
      ||||| : : : : : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy      65 IAAPIIDAISLQGVNGTIFAYGQTASGKTYTMMG----SEDLGVIPRAIHDIPOKIKKFP 120
      ||||| : : : : : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db      61 VAHPIVDDCFHGYNCITFAYGQTGSKTYTMMGPGGEHPDHMGIIIPCCHDIFRIDKQF 120
      ||||| : : : : : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy      121 DRE---FLLRVSYMEIINVEITDILLC-GTQMKPLIITREDVNRNVYADLTETVVYVYSEM 176
      ||| : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db      121 EKQHDFFHWKCSYMEIINVEIYDILLCPNQHPQLNIHEHPNMGPPYQGGCTFEHVCSYED 180
      ||||| : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy      177 ALKWITGKSRVAGETKMNQRSRSHITFIRMILESEKGPSCGSKVYSHLNLVDLA 236
      ||||| : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db      181 ACHWIOGNNRHHVAATNMDDHSRSHITFIHVQQRHK----QCDEHVCHSKNVLVDLA 236
      ||||| : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy      237 GSRAAQTGAAGVRLKEGCGNINRSLFILGOVYKKLSDGQV-----GGFINVRSKLTRI 290
      ||||| : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db      237 GSERVNETGAEGQLKEGCGNINQSLTTLGNVINALADGQTKYMTGGHGHIPIYRDSKLTWL 296
      ||||| : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy      291 LQNSLGNPKTRITCTITPV--SDEPILTALQFASTAKYMKNTYVNE 336
      ||||| : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db      297 LQDSLGNCKCTMLACIWPADWNYEETLSTLYADRAKNIKKPKQINE 344
      ||||| : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

```

RESULT 6	
ABU53125	ABU53125 standard; protein; 366 AA.
ID	ABU53125 standard; protein; 366 AA.
XX	
AC	ABU53125;
XX	
DT	15-APR-2003 (first entry)
XX	
DE	Intracellular trafficking-associated DKFZphtes3_26g22 homologue #2.
XX	
KW	Human; gene therapy; vaccine; disease treatment; detection.
XX	
OS	Homo sapiens.
XX	
PN	WO200112659-A2.
XX	
PD	22-FEB-2001.
XX	
PF	18-AUG-2000; 2000WO-IB001496.
XX	
PR	18-AUG-1999; 99US-0149499P.
PR	28-SEP-1999; 99US-0156503P.
XX	
PA	(GEHU-) GERMAN HUMAN GENOME PROJECT.
XX	
PI	Wiemann S;
XX	
DR	WPI; 2001-327840/34.
XX	
PT	Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies.
XX	
PS	Example III; Page 745; 1095pp; English.
XX	
CC	This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the

CC presence of polypeptides in samples. This sequence represents a homologue
CC of a polypeptide described in the disclosure of the invention
XX
SQ Sequence 366 AA;

Query Match 44.8%; Score 781; DB 4; Length 366;
Best Local Similarity 51.1%; Pred. No. 7.1e-68;
Matches 178; Conservative 38; Mismatches 106; Indels 26; Gaps 7;

QY 11 RVRPLNRESLGETAQVYK-----KTDNNVIYQVDSKSFNDFRVFHNQETTKNYYE 64
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1 RCRPLNERINDGSCCVQWPPWGTGKTVHNGHEGSPKSFDFHFWNCTQEDVYDT 60
QY 65 IAAPILDSAIQYNGTIFAYGQTASGKTYTMMG-----SEDLGLVPIRAIHDIPOKIKKFP 120
Db : ||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 61 VAHPVDDCFHGYNCITIFAYGQTGSKTYTMMGPGGHEPDHMGIIIPCCCHDIPDRIDKFO 120
Db : ||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 121 DRE---FLRVSYMEIYNETITDILC-GTQMKPLIREDVNRNYYVADLVEEYVYSEM 176
Db : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 121 EKDHDFHWKCSYMEIYNEEYDILCPNPQHKPLNIHHPNMPYVQCTEFHVCSEYD 180
QY 177 ALKWTGKSRHYGKTKNQSRSSHTIFRMILESEKGEPSNCEGSKVYSHLNLVDLA 236
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 181 ACHWQGNKRVAAATNNDSSRSHTIFTIHEQRHK---QCDEHYCHSKMNLVDLA 236
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 237 GSERAAQTGAAGVRLKEGNCINRSFILGOVIKLSDGQV-----GGFINYRDSKLTRI 290
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 237 GSERNRTGAEORLKEGNCINQSLTLGNVINALADGQTKYMGHGHIPIYRDSKLTL 296
QY 291 LQNSLGGNPKTRIICITPV--SFDETLALQFASHTAKYMKNTPYNE 336
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 297 LQDSLGGNCKTMIACIWPADWNBYETLSTLYADRAKNIKNKPOINE 344
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
RESULT 7
ABB62322
ID ABB62322 standard; protein; 2013 AA.
XX
AC ABB62322;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 13758.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
FN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
ER 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL06425.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 13758; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2013 AA;

Query Match 40.5%; Score 707; DB 4; Length 2013;
Best Local Similarity 49.4%; Pred. No. 1.8e-59;
Matches 166; Conservative 51; Mismatches 103; Indels 16; Gaps 8;

QY 5 AVAVCVRVPLNRESLGETAQVYKWT-DNNVIYQVDSKSFNDFRVFHNQETTKNYY 62
Db : ||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 8 SIQVCIKVRPCPEGLTSL-----WQVKERSIHLADSHAEPYVDFYVDEGASNQEVF 60
QY 63 BEIAAPIIDSAIQYNGTIFAYGQTASGKTYTMMGSEDLGLVPIRAIHDIPOKIKKFPDR 122
Db : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 61 DMWAKHIVHACQGFNGTIFAYGQTSSGKTYTMMGDEQNPQGVNLAKEIFQOISSETER 120
QY 123 EELLRVSYMEIYNETITDILCCTQMKPLIREDVNRNYYVADLVEEYVYSEM-ALKWI 181
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 121 DFLAVGVIEIYNEKIYDILL--NKNQDLKHESNGIVNVN--CECIIITSEVDLLRL 176
QY 182 TGKESRHYGKTKNQSRSSHTIFRMILESEKGEPSNCEGSKVYSHLNLVDLAGSERA 241
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 177 CLGNKERTVGETNMNERSRSRSHAIKFIIESR-KSDHSD-DDAVIQSVNLVDLAGSERA 234
QY 242 AQTGAAGVRLKEGNCINRSFILGOVIKLSDGQVGGFINYRDSKLTRIQLNSLGGNPKT 301
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 235 DOTGARGARLKEGGHINKSLFLSNVIKLSNADNRFTNYRDSKLTRIQLNSLGGNPKT 294
QY 302 RIICITPVSFDETLALQFASHTAKYMKNTPYNEV 337
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 295 SIICITKESIMEESQSTLSFATRAKKIRIKPOVNEV 330
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
RESULT 8
ABB70992
ID ABB70992 standard; protein; 473 AA.
XX
AC ABB70992;
XX
DT 10-DEC-2002 (first entry)
XX
DE Human target protein.
XX
KW Microtubule stimulated ATPase; cellular proliferation; cancer; AIDS;
KW hyperplasia; restenosis; cardiac hypertrophy; immune disorder; HIV;
KW inflammation; autoimmune disease; arthritis; graft rejection; psoriasis;
KW inflammatory bowel disease; proliferation; medical procedure; surgery;
KW human immunodeficiency virus; acquired immunodeficiency syndrome;
KW angioplasty; human; Hskif; kinesin family.
XX
OS Homo sapiens.
XX
FN US6440684-B1.
XX
PD 27-AUG-2002.
XX
PF 12-JUN-2000; 2000US-00592054.
XX
PR 12-JUN-2000; 2000US-00592054.
XX
PA (CYTO-) CYTOKINETICS INC.
XX
PI Beraud C, Finer JT, Sakowicz R, Wood KW;
XX
DR WPI; 2002-711529/77.
DR N-PSDB; ABB55162.

XX PD 06-MAR-2003.

XX PF 23-AUG-2002; 2002WO-GB003892.

XX PR 23-AUG-2001; 2001GB-00020558.

XX PR 05-OCT-2001; 2001GB-00024037.

XX PA (OXFO-) OXFORD BIOMEDICA UK LTD.

XX PI Kingsman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;

XX PD WPI; 2003-290046/28.

XX DR N-PSDB; ADD18925.

XX XX New substantially purified polypeptide, useful for diagnosing or treating a hypoxia-regulated condition, such as cancer, ischemia, reperfusion injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or wound healing.

XX PS Claim 25; SEQ ID NO 413; 424pp; English.

XX XX This invention relates to novel human genes and gene product which are implicated in certain disease states. Compounds which modulate the proteins of the invention may have cytostatic, antiinflammatory, ophthalmological, antiarteriosclerotic or vulnerary activities. The sequences of the invention may be useful for gene therapy. The invention may be useful for diagnosing or treating a hypoxia-regulated condition, such as tumorigenesis, angiogenesis, apoptosis, inflammation, erythropoiesis, or the biological response to hypoxia conditions including processes such as glycolysis, gluconeogenesis, glucose transportation, catecholamine synthesis, iron transport or nitric oxide synthesis. The disease includes cancer, ischaemic conditions, reperfusion injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis, CC inflammatory conditions or wound healing. The present sequence is that of a disease related protein of the invention.

XX XX Sequence 1232 AA;

Query Match 38.3%; Score 668; DB 7; Length 1232;

Best Local Similarity 44.7%; Pred. No. 5.9e-56;

Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 5;

QY 6 VAVCVRLPLNSREELG-ETAQVYVTKDNNVIYQVDSKSFNDRVPHGNKTNVYEE 64

Db 10 VRVALRCEPLVPKEISECCMLSFVGPQVVGTD--KSFTYDFVDPSTEQBEVNT 67

QY 65 IAAPIIDSAIOGYGTIFAYQGTASGKTYTWG-----SEDLGVIPRAIHIFQKIK 117

Db 68 AVAPLIKGVFGYNATVLAYGQTSGKYSMGAYTAQEENPTVGVIPRVLIQLFKRID 127

QY 118 KPDREFLLRVSYMEIYNETITDLCTQKMKPLIREDVNRNVYVADLTREVVYVTSMA 177

Db 128 KKSDFEFTLRKVSYLEIYNEIRILDLCPSREKAQINIREDPKEGKIVGLTERTVULVADLT 187

QY 178 LKWTIKGKSHYGTKNORSRSHITFRMILSRKGEPSNCEGSKVSHLNLVDIAG 237

Db 188 VSCLEQGNNSRTVASTAMNSQSRSHAFTISLEQRKSD---KNSSFRSKLHLVDLAG 243

QY 238 SERAAQTGAAGVRLKEGNNINSLFTLGQVINKSLSDGOVGFINTRDSKLTILQNSLGG 297

Db 244 SERQKTKAEGDRLEKGININRGLLCLGNVISALGDDKGGFVPRYDRSKLTRELLQDSLGG 303

QY 298 NPKTILICTIPV--SFDETLALQFASAKYMKNTPVN 335

Db 304 NSHTLMIACVSPADSNLBETLNLRYADARKIKNKPVN 343

RESULT 12

ADD4937

ID ADD4937 standard; protein; 1232 AA.

XX AC ADD49937;

XX DT 15-JAN-2004 (first entry)

XX DE Human kinesin family member 4A.

XX XW Human; lung cancer antigen; cytostatic; lung cancer; gene therapy; vaccine; T-cell; tumour.

XX OS Homo sapiens.

XX PN US2003194764-A1.

XX PD 16-OCT-2003.

XX PF 04-APR-2002; 2002US-00116712.

XX PR 05-APR-2001; 2001US-0282289P.

XX PR 03-OCT-2001; 2001US-0327511P.

XX PA (CORI-) CORIXA CORP.

XX PI Bangur CS, Switzer A.

XX DR WPI; 2003-844452/78.

XX DR N-PSDB; ADD49935.

XX PT New isolated polypeptides and polynucleotides useful for diagnosing, preventing and treating cancer, particularly lung cancer.

XX PS Example 3; SEQ ID NO 669; 250pp; English.

XX XX The invention relates to an isolated polynucleotide (a) comprising any of the 666 fully defined nucleotide sequences appearing as ADD49289 at ADD49935, ADD49936 and ADD49938, complements of (a); sequences of at least 20 contiguous residues of (a); sequences that hybridise to (a) under highly stringent conditions; sequences having at least 75 or 90% identity to (a); or degenerate variants of (a). Also included are an isolated polypeptide (b) (comprising: sequences encoded by the new polynucleotide; any of the 4 amino acid sequences fully defined in the specification; or sequences having at least 70 or 90% identity to the sequence in (a) or (b)), an expression vector comprising the above polynucleotide operably linked to an expression control sequence, a host cell transformed or transfected with the above expression vector, an isolated antibody, or its antigen-binding fragment, that specifically binds to the above polypeptide, an oligonucleotide that hybridises to the above-mentioned nucleotide sequences under highly stringent conditions, a fusion protein comprising at least one polypeptide cited above, detecting the presence of a cancer in a patient (comprising: obtaining a biological sample from the patient; contacting the biological sample with a binding agent that binds to the polypeptide, or with the oligonucleotide cited above; detecting in the sample an amount of the polypeptide that binds to the binding agent, or an amount of a polynucleotide that hybridises to the oligonucleotide; and comparing the amount of polypeptide, or polynucleotide that hybridises to the oligonucleotide, to a predetermined cut-off value and then determining the presence of a cancer in the patient), a method for stimulating and/or expanding T-cells specific for a tumour protein (comprising: contacting T-cells with the above polypeptide, polynucleotide or antigen-presenting cells that express the polynucleotide, under conditions and for a time sufficient to permit the stimulation and/or expansion of T-cells), an isolated T-cell population comprising T-cells prepared by the method, a composition comprising a first component selected from physiological carriers and immunostimulants, and a second component selected from the above polypeptide, polynucleotide, antibody, fusion protein, T-cell population and antigen-presenting cells that express the above polypeptide, stimulating an immune response in a patient (comprising administering to the patient the above composition) treating lung cancer in a patient (comprising administering to the patient the above composition and a diagnostic kit (comprising: at least one oligonucleotide cited above; or at least one antibody cited above and a detection reagent, where the detection reagent comprises a reporter group). The composition and methods are useful in diagnosing, preventing and treating cancer, particularly lung cancer. The present sequence is a lung cancer-

CC associated antigen of the invention.

XX Sequence 1232 AA;

Query Match 37.7%; Score 658; DB 7; Length 1232;
Best Local Similarity 44.7%; Pred. No. 5.6e-55;
Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 6;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDFRVFHNKNTKNVYEE 64
DQ 10 VRVALRCRPLVPKEISEGCMCLSFVPGEPQVVVGTD--KSTYDFVDFPSTQEVEFNT 67
QY 65 IAAPIDSAIQYNGTIFAYGQTASGKTYTMG-----SEHLGVIPRAIHDPQKIK 117
DQ 68 AVAPLKGKGVKYNATVLAQGTGSGKTYSMGGAATAEQNEPTVGVIPRVQLLFKEID 127
QY 118 KFPDFEFLRVSYMEIYNETITDLCGTOKMKPLIREDVNRNRYVADLTVEVVYVTEMA 177
DQ 128 KKSDFEFLRVSYMEIYNEEILDLCPSEKAQINREDPKEGKIVGTEKTVLVALDT 187
QY 178 LKWTGKESRHYGETKMNQSRSHITFRMILESEKGEPCNCEGSKVSHNLVDLAG 237
DQ 188 VSCLEQGNNSRTVASTAMNSQSSRSHAFTI---SLEQKKSKDNSSFR-SKLHLVDLAG 243
QY 238 SERAAQTGAAGVRLKEGNCNINRSFLIGQVKKLSDGQGVGFYNYRDSKLTILQNSLGG 297
DQ 244 SERQKTKAEGDRLKEGINRGLLCLGNVISALGDDKKGFPYRDSKLTILQNSLGG 303
QY 298 NPKTRIICTIPV--SPDETITLQFASTAKYMKNTPYVN 335
DQ 304 NSHTLMIACVSPADSNLEETLTLRYADRAKRNKPIVN 343

RESULT 13
ADD49932

ID ADD49932 standard; protein; 1232 AA.

AC ADD49932;

DT 15-JAN-2004 (first entry)

XX Human lung specific tumour antigen L 1477p.

XX Human; lung cancer antigen; cytostatic; lung cancer; gene therapy;
XX vaccine; T-cell; tumour.

OS Homo sapiens.

PN US2003194764-A1.

PD 16-OCT-2003.

XX 04-APR-2002; 2002US-00116712.

XX 05-APR-2001; 2001US-0282289P.

XX 05-OCT-2001; 2001US-0327511P.

XX (CORI-) CORIXA CORP.

XX Bangor CS, Switzer A;

XX WPI; 2003-844452/78.

XX N-PSDB; ADD49929.

XX New isolated polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cancer, particularly lung cancer.

XX Example 3; SEQ ID NO 664; 250pp; English.

XX The invention relates to an isolated polynucleotide (a) comprising any of
XX the 666 fully defined nucleotide sequences appearing as ADD49269 -
XX ADD49935, ADD49936 and ADD49938, complements of (a); sequences of (a)
XX at least 20 contiguous residues of (a); sequences that hybridise to (a)

CC under highly stringent conditions; sequences having at least 75 or 90%
CC identity to (a); or degenerate variants of (a). Also included are an
CC isolated polypeptide (b) comprising: sequences encoded by the new
CC polynucleotide; any of the 4 amino acid sequences fully defined in the
CC specification; or sequences having at least 70 or 90% identity to the
CC sequence in (a) or (b), an expression vector comprising the above
CC polynucleotide operably linked to an expression control sequence, a host
CC cell transformed or transfected with the above expression vector, a host
CC isolated antibody, or its antigen-binding fragment, that specifically
CC binds to the above polypeptide, an oligonucleotide that hybridises to the
CC above-mentioned nucleotide sequences under highly stringent conditions, a
CC fusion protein comprising at least one polypeptide cited above, detecting
CC the presence of a cancer in a patient (comprising: obtaining a biological
CC sample from the patient; contacting the biological sample with a binding
CC agent that binds to the polypeptide, or with the oligonucleotide cited
CC above; detecting in the sample an amount of the polypeptide that binds to
CC the binding agent, or an amount of a polynucleotide that hybridises to
CC the oligonucleotide; and comparing the amount of polypeptide, or
CC polynucleotide that hybridises to the oligonucleotide, to a predetermined
CC cut-off value and then determining the presence of a cancer in the
CC patient), a method for stimulating and/or expanding T-cells specific for
CC a tumour protein (comprising contacting T-cells with the above
CC polypeptide, polynucleotide or antigen-presenting cells that express the
CC polynucleotide, under conditions and for a time sufficient to permit the
CC stimulation and/or expansion of T-cells), an isolated T-cell population
CC comprising T-cells prepared by the method, a composition comprising a
CC first component selected from physiological carriers and
CC immunostimulants, and a second component selected from the above
CC polypeptide, polynucleotide, antibody, fusion protein, T-cell population
CC and antigen-presenting cells that express the above polypeptide,
CC stimulating an immune response in a patient (comprising administering to
CC the patient the above composition) treating lung cancer in a patient
CC (comprising administering to the patient the above composition and a
CC diagnostic kit (comprising: at least one oligonucleotide cited above; or
CC at least one antibody cited above and a detection reagent, where the
CC detection reagent comprises a reporter group). The composition and
CC methods are useful in diagnosing, preventing and treating cancer,
CC particularly lung cancer. The present sequence is a lung cancer-
CC associated antigen of the invention.

XX Sequence 1232 AA;

Query Match 37.7%; Score 658; DB 7; Length 1232;

Best Local Similarity 44.7%; Pred. No. 5.6e-55;

Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 6;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDFRVFHNKNTKNVYEE 64
DQ 10 VRVALRCRPLVPKEISEGCMCLSFVPGEPQVVVGTD--KSTYDFVDFPSTQEVEFNT 67
QY 65 IAAPIDSAIQYNGTIFAYGQTASGKTYTMG-----SEHLGVIPRAIHDPQKIK 117
DQ 68 AVAPLKGKGVKYNATVLAQGTGSGKTYSMGGAATAEQNEPTVGVIPRVQLLFKEID 127
QY 118 KFPDFEFLRVSYMEIYNETITDLCGTOKMKPLIREDVNRNRYVADLTVEVVYVTEMA 177
DQ 128 KKSDFEFLRVSYMEIYNEEILDLCPSEKAQINREDPKEGKIVGTEKTVLVALDT 187
QY 178 LKWTGKESRHYGETKMNQSRSHITFRMILESEKGEPCNCEGSKVSHNLVDLAG 237
DQ 188 VSCLEQGNNSRTVASTAMNSQSSRSHAFTI---SLEQKKSKDNSSFR-SKLHLVDLAG 243
QY 238 SERAAQTGAAGVRLKEGNCNINRSFLIGQVKKLSDGQGVGFYNYRDSKLTILQNSLGG 297
DQ 244 SERQKTKAEGDRLKEGINRGLLCLGNVISALGDDKKGFPYRDSKLTILQNSLGG 303
QY 298 NPKTRIICTIPV--SPDETITLQFASTAKYMKNTPYVN 335
DQ 304 NSHTLMIACVSPADSNLEETLTLRYADRAKRNKPIVN 343

RESULT 14
ABG70990

ABG70990 standard; protein; 1232 AA.
ABG70990;
10-DEC-2002 (first entry)
Human HsKif4 protein.
Microtubule stimulated ATPase; cellular proliferation; cancer; AIDS;
hyperplasia; restenosis; cardiac hypertrophy; immune disorder; HIV;
inflammation; autoimmune disease; arthritis; graft rejection; psoriasis;
inflammatory bowel disease; proliferation; medical procedure; surgery;
human immunodeficiency virus; acquired immunodeficiency syndrome;
angioplasty; human; HsKif4; kinesin family.
Homo sapiens.
Key Location/Qualifiers
Misc-difference 161..162 /note= "Encoded by TTAGAG"
Misc-difference 164..180 /note= "Encoded by TACAATGAAGAAATTGGATCTTCTAATGTC-CATCTCGTGAGAAAGCTCAA"
US6440684-B1.
27-AUG-2002.
12-JUN-2000; 2000US-00592054.
12-JUN-2000; 2000US-00592054.
(CYTO-) CYTOKINETICS INC.
Beraud C, Finer JT, Sakowicz R, Wood KW;
WPI; 2002-711529/77.
N-PSDB; ABS55160.
Screening for modulators of target protein having microtubule stimulated ATPase activity e.g. kinesin family of protein, useful for treating cancer, psoriasis, arthritis, human immunodeficiency virus (HIV) infection.
Claim 1; Fig 2; 34pp; English.
The present invention relates to a new method of screening modulators of target protein with microtubule stimulated ATPase activity. The method involves contacting the target protein with an agent at 1st and 2nd concentrations and determining the level of activity (e.g. binding or ATPase activity) of target protein, where a difference between levels of activity of target protein contacted with 1st and 2nd concentrations of an agent indicates that an agent modulates activity of target protein. The invention can be used for screening for modulators of target protein having microtubule stimulated ATPase activity. The compounds identified by method of the invention are useful for treating cellular proliferation including cancer, hyperplasias, restenosis, cardiac hypertrophy, immune disorders and inflammation. The compounds identified by the method are also useful for treating autoimmune disease, arthritis, graft rejection, inflammatory bowel disease, proliferation induced by medical procedures, e.g. surgery, angioplasty etc. The compounds are also useful for treating psoriasis. The compounds are useful for inhibiting human immunodeficiency virus (HIV) and thus treating acquired immunodeficiency syndrome (AIDS). The present amino acid sequence represents the human HsKif4 (kinesin family) protein of the invention
Sequence 1232 AA;
Query Match 37.4%; Score 652; DB 5; Length 1232;
Best Local Similarity 44.4%; Pred. No. 2.2e-54;
Matches 151; Conservative 52; Mismatches 121; Indels 16; Gaps 6;
6 VAVCVRFLNGBESLG-ETAQVYWKTDNNVIYQVDSKSFNDRVPHGNETTKNVEE 64

Db 10 VVALRCRLVPKEISEGQCMCLSFVGPQVVGTD--KSTYDFVDPSTEQBEVENT 67
QY 65 IAAPIIDSAIOGYNGTIFAYGQTASGKTYTMWG-----SEHLGVIPRAHDIPQKIK 117
Db 68 AVAPLIKGVKGYNATVLAAYGQTSGKTYSMGAYTAEQENEPTVGVIPIRVQLLFKEID 127
QY 118 KPDRREFLURVSYMEIYNETITDLGCTQKMKPLIITREDVNNVYVADLTBEVVYVYSEMA 177
Db 128 KKSDEFTLKVSYLEIYNEEIIIDLCPSREKAQINIREDPKEGKIVGLTEKTVLVALDT 187
QY 178 LKMWTKGKSRHYGETKMNORSRSHITIFRMILESEKGEPSNCEGSVKVSHNLVLDLAG 237
Db 188 VSCLEGNNRSTVASTAMNSQSSRAILTI---SLEQKKSKDKNSSFR-SKLHLVLDLAG 243
QY 238 SERAAQTGAAGVRLKEGCNINRSLFILGQVIKLSGQVGGFNYNEDSKLITLQNSLGG 297
Db 244 SERQKTKAEGDRLKEGINIRGLLCGNVISALGDDKGGFAPYRDSKLTLLQDSLGG 303
QY 298 NPKTRILICTIPV--SFDETLTALQFASTAKYMNTPYVN 335
Db 304 NSHTLMACVSPADSNLEETLNLTRYADRARKINKPIVN 343
RESULT 15
ABG70993
ID ABG70993 standard; protein; 1234 AA.
XX
AC ABG70993;
XX
DT 10-DEC-2002 (first entry)
XX
DE Human HsKif4b protein.
XX
KW Microtubule stimulated ATPase; cellular proliferation; cancer; AIDS;
hyperplasia; restenosis; cardiac hypertrophy; immune disorder; HIV;
inflammation; autoimmune disease; arthritis; graft rejection; psoriasis;
inflammatory bowel disease; proliferation; medical procedure; surgery;
human immunodeficiency virus; acquired immunodeficiency syndrome;
angioplasty; human; HsKif4b; kinesin family.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 850 /note= "Encoded by CGBCA"
FT FT
XX
XX US6440684-B1.
XX
PD 27-AUG-2002.
XX
PF 12-JUN-2000; 2000US-00592054.
XX
PR 12-JUN-2000; 2000US-00592054.
XX
PA (CYTO-) CYTOKINETICS INC.
XX
XX Beraud C, Finer JT, Sakowicz R, Wood KW;
XX
XX WPI; 2002-711529/77.
XX
XX N-PSDB; ABS55163.
XX
XX Screening for modulators of target protein having microtubule stimulated ATPase activity e.g. kinesin family of protein, useful for treating cancer, psoriasis, arthritis, human immunodeficiency virus (HIV) infection.
XX
XX Disclosure; Fig 8; 34pp; English.
XX
XX The present invention relates to a new method of screening modulators of target protein with microtubule stimulated ATPase activity. The method involves contacting the target protein with an agent at 1st and 2nd concentrations and determining the level of activity (e.g. binding or ATPase activity) of target protein, where a difference between levels of activity of target protein contacted with 1st and 2nd concentrations of an agent indicates that an agent modulates activity of target protein. The invention can be used for screening for modulators of target protein having microtubule stimulated ATPase activity. The compounds identified by method of the invention are useful for treating cellular proliferation including cancer, hyperplasias, restenosis, cardiac hypertrophy, immune disorders and inflammation. The compounds identified by the method are also useful for treating autoimmune disease, arthritis, graft rejection, inflammatory bowel disease, proliferation induced by medical procedures, e.g. surgery, angioplasty etc. The compounds are also useful for treating psoriasis. The compounds are useful for inhibiting human immunodeficiency virus (HIV) and thus treating acquired immunodeficiency syndrome (AIDS). The present amino acid sequence represents the human HsKif4 (kinesin family) protein of the invention
Sequence 1232 AA;
Query Match 37.4%; Score 652; DB 5; Length 1232;
Best Local Similarity 44.4%; Pred. No. 2.2e-54;
Matches 151; Conservative 52; Mismatches 121; Indels 16; Gaps 6;
6 VAVCVRFLNGBESLG-ETAQVYWKTDNNVIYQVDSKSFNDRVPHGNETTKNVEE 64

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:26:01 ; Search time 160.292 Seconds
(without alignments)
4694.096 Million cell updates/sec

Title: US-10-045-631B-88

Perfect score: 13329

Sequence: 1 MAEGAVAVCVRRVPLNSRE.....SQPGMHASSGKDVPCKTQ 2663

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_29Jan04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13305	99.8	2663	4 AAM39097	Aam39097 Human pol
2	13170	98.8	2688	4 AAM40883	Aam40883 Human pol
3	13131	98.5	2633	4 ABG06505	Abg06505 Novel hum
4	13115	27.1	2954	2 AAY01632	Aay01632 Amino aci
5	1545.5	11.6	2013	4 ABB62322	Abb62322 Drosophil
6	1286	9.6	1931	4 ABB61012	Abb61012 Drosophil
7	1183	8.9	240	5 AAU83007	Aau83007 Human hom
8	1127.5	8.5	3259	7 ADE56037	Ades6037 Human pro
9	1127.5	8.5	3259	7 ADE56033	Ades6033 Human pro
10	1091	8.2	2228	7 ABR61599	Abr61599 Human gol
11	1091	8.2	2230	6 ABU07445	Abu07445 Protein d
12	1091	8.2	2230	7 ABR61600	Abr61600 Human gol
13	1089	8.2	2252	7 ABR61602	Abr61602 Human gol
14	1089	8.2	3187	7 ADE56031	Ades6031 Rat Prote
15	1089	8.2	3187	7 ADE56035	Ades6035 Rat Prote
16	1089	8.2	3187	7 ADE56035	Ades6035 Rat Prote
17	1077.5	8.1	1388	5 AAU14400	Aau14400 Human kin
18	1077.5	8.1	1388	5 AAU79590	Aau79590 Human kin
19	1077.5	8.1	1388	6 ABR48222	Abr48222 Human bla
20	1077.5	8.1	1388	6 ADC35116	Adc35116 Human bre
21	1072.5	8.0	1388	7 ADB80468	Adb80468 Ovarian c
22	1064	8.0	3899	6 ABR92048	Abr92048 Human cer
23	1064	8.0	3917	6 ABR92050	Abr92050 Human cer
24	1058	7.9	3907	6 ABR92047	Abr92047 Human cer
25	1058	7.9	3925	6 ABR92049	Abr92049 Human cer

ALIGNMENTS

RESULT 1

AAM39097
ID AAM39097 standard; protein; 2663 AA.

XX AC AAM39097;

DT 22-OCT-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 2242.

XX KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
leukaemia.

OS Homo sapiens.

XX KW WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US034263.

XX PR 23-DEC-1999; 99US-00471275.

PR 21-JAN-2000; 2000US-00488725.

PR 25-APR-2000; 2000US-0052317.

PR 20-JUN-2000; 2000US-00598042.

PR 19-JUL-2000; 2000US-00620312.

PR 03-AUG-2000; 2000US-00653450.

PR 14-SEP-2000; 2000US-00662191.

PR 19-OCT-2000; 2000US-00693036.

PR 29-NOV-2000; 2000US-00727344.

(HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

XX N-PSDB; AAI58253.

PT Novel nucleic acids and polypeptides, useful for treating disorders such
as central nervous system injuries.

XX Example 4; SEQ ID NO 2242; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and the
CC encoded polypeptides (AAM3642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
SQ Sequence 2663 AA;
Query Match 99.8%; Score 13305; DB 4; Length 2663;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2659; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 MAEGAVAVCVVRPLNRSREESLGETAQVYKTDNNVIYQVDSKSFNDRVFHGNETK 60
DB 1 MAEGAVAVCVVRPLNRSREESLGETAQVYKTDNNVIYQVDSKSFNDRVFHGNETK 60
QY 61 NYVEETAAPIIDSAIQGYNGTIFAYGQTASGKTYTMMGSEDLGVIPRAIHDFQKIKKF 120
DB 61 NYVEETAAPIIDSAIQGYNGTIFAYGQTASGKTYTMMGSEDLGVIPRAIHDFQKIKKF 120
QY 121 PDREFLLRVSYMEIYNETITDILCGTQKMKPLIIRREDVNRNVYVADLTSEVVYVSEMALK 180
DB 121 PDREFLLRVSYMEIYNETITDILCGTQKMKPLIIRREDVNRNVYVADLTSEVVYVSEMALK 180
QY 181 WITKGEKSHYGETKMNQSRSSHTIFRMILSREKGEPSNCEGSYKVSMLNLDVLAGSE 240
DB 181 WITKGEKSHYGETKMNQSRSSHTIFRMILSREKGEPSNCEGSYKVSMLNLDVLAGSE 240
QY 241 RAAQTGAAGVRLKEGGINRSFILGOVVKLSDGQGVGFIVNRDSKLTFRILQNSLGGNP 300
DB 241 RAAQTGAAGVRLKEGGINRSFILGOVVKLSDGQGVGFIVNRDSKLTFRILQNSLGGNP 300
QY 301 KTRIICTITPVSFDETLTALQFASHTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKOLE 360
DB 301 KTRIICTITPVSFDETLTALQFASHTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKOLE 360
QY 361 EVSLETRAQAMEKDQALLEEKDLLOKVQNEKIENLTMIVTSSSLTQQELKAKRKR 420
DB 361 EVSLETRAQAMEKDQALLEEKDLLOKVQNEKIENLTMIVTSSSLTQQELKAKRKR 420
QY 421 VTWCLGKINKMNSYADQFNIPNTITTKHLSINLLREIDRSVCSESDFVFNLTDLTSL 480
DB 421 VTWCLGKINKMNSYADQFNIPNTITTKHLSINLLREIDRSVCSESDFVFNLTDLTSL 480
QY 481 EIEWNPATKLLNQENIESELSLRADYDNLVLDYEQRLTEKEEMELKKEKNLDLFEAL 540
DB 481 EIEWNPATKLLNQENIESELSLRADYDNLVLDYEQRLTEKEEMELKKEKNLDLFEAL 540
QY 541 ERKTKKQDMQLTHEISNLKLVKRVHVNQDLLENELSSKVLEIRKEKDOIKKLOEYIDS 600
DB 541 ERKTKKQDMQLTHEISNLKLVKRVHVNQDLLENELSSKVLEIRKEKDOIKKLOEYIDS 600
QY 601 QKLENTKMDLSYLSIESIEDPKQKOTLFDATVALDAKRESAFILSENLELKEKMKELAT 660
DB 601 QKLENTKMDLSYLSIESIEDPKQKOTLFDATVALDAKRESAFILSENLELKEKMKELAT 660
QY 661 TYQKQENDIQLYQSLEAKKQVVDLEKELQSAFNEITKLTSLIDGVKPKDLCNLEBEG 720
DB 661 TYQKQENDIQLYQSLEAKKQVVDLEKELQSAFNEITKLTSLIDGVKPKDLCNLEBEG 720
QY 721 KITDLOKELNKEVEENALREEVILLSELKSPSEVERLRKEIQDKSEELHIITSEKDKL 780

DB 721 KITDLOKELNKEVEENALREEVILLSELKSPSEVERLRKEIQDKSEELHIITSEKDKL 780
QY 781 FSEVVHRESRVQGLLEIGIKTKDLDLATTSQSNKSTDOEFQNFKTLHMDFEQYKMWLEEN 840
DB 781 FSEVVHRESRVQGLLEIGIKTKDLDLATTSQSNKSTDOEFQNFKTLHMDFEQYKMWLEEN 840
QY 841 ERMQEIVNLSKEAQKPFSSLGALKTSLSYTQELQEKTRVQERLNEMEQLEKLENRD 900
DB 841 ERMQEIVNLSKEAQKPFSSLGALKTSLSYTQELQEKTRVQERLNEMEQLEKLENRD 900
QY 901 SPLQTVREKTLITTEKLOOTLEEVKTLTQEKDDLKQLOESIQIERDQKSDIHDVTNNVI 960
DB 901 SPLQTVREKTLITTEKLOOTLEEVKTLTQEKDDLKQLOESIQIERDQKSDIHDVTNNVI 960
QY 961 DTQEOLRNALSLKQHQFTINTLKSISEEVSRLNHEMNETGETKDFQCKWVGDKKQD 1020
DB 961 DTQEOLRNALSLKQHQFTINTLKSISEEVSRLNHEMNETGETKDFQCKWVGDKKQD 1020
QY 1021 LEAKNTQTLTADVKNNEIIEQORKIFSLIQEKNELOQMLESVIAEKEQKTDLKENIEMT 1080
DB 1021 LEAKNTQTLTADVKNNEIIEQORKIFSLIQEKNELOQMLESVIAEKEQKTDLKENIEMT 1080
QY 1081 IENQBELRLGDELKKQOEIQAQEKHAIKKEGELSRCTCDRLAARVEEKLKEKSOQLOEK 1140
DB 1081 IENQBELRLGDELKKQOEIQAQEKHAIKKEGELSRCTCDRLAARVEEKLKEKSOQLOEK 1140
QY 1141 QOLLNVQEMSEMQKINEIENLKNELKNKELTLEHMETERLELAQKLNENVEEVKSITK 1200
DB 1141 QOLLNVQEMSEMQKINEIENLKNELKNKELTLEHMETERLELAQKLNENVEEVKSITK 1200
QY 1201 ERKVLKELQKSPETBRDLRGVIRIEATGLQTEBELKIAHILHKEHQFTIDELRRSVSE 1260
DB 1201 ERKVLKELQKSPETBRDLRGVIRIEATGLQTEBELKIAHILHKEHQFTIDELRRSVSE 1260
QY 1261 KTAQIINTQDLEKSHTKLOEIPVLHBEQELLPNVKVSETOETMNELELLTEQSTTKDS 1320
DB 1261 KTAQIINTQDLEKSHTKLOEIPVLHBEQELLPNVKVSETOETMNELELLTEQSTTKDS 1320
QY 1321 TTLARIEMRLRLNKEFQBSQOEIISLTKERONLAKITKEALEVKHDLKEHIRETLAKIQ 1380
DB 1321 TTLARIEMRLRLNKEFQBSQOEIISLTKERONLAKITKEALEVKHDLKEHIRETLAKIQ 1380
QY 1381 ESQSQEQSLNKEKONNETTKIVSEMEQPKDSALLRIEIMGLSKLOESHDMSKV 1440
DB 1381 ESQSQEQSLNKEKONNETTKIVSEMEQPKDSALLRIEIMGLSKLOESHDMSKV 1440
QY 1441 AKEKDQLQRLQVLSQESDQKLENIKEIVAKHLETEBELKVAHCCLKEQEBETINELRVNL 1500
DB 1441 AKEKDQLQRLQVLSQESDQKLENIKEIVAKHLETEBELKVAHCCLKEQEBETINELRVNL 1500
QY 1501 SEKETIESTIQKLEAINDKLQNKIQEIVYEKEBOLNIKOISEVOENVNELKQFKEHRKAK 1560
DB 1501 SEKETIESTIQKLEAINDKLQNKIQEIVYEKEBOLNIKOISEVOENVNELKQFKEHRKAK 1560
QY 1561 DSALQIESKMLLETNRLQESQOEIQIMIKEEMKRVQEQALQIERDQKENTKEIVAKM 1620
DB 1561 DSALQIESKMLLETNRLQESQOEIQIMIKEEMKRVQEQALQIERDQKENTKEIVAKM 1620
QY 1621 KESQEKYQFLKMTAVNETQEKMCETIEHLKQBPETQKLNLENETENIRLTQILHENLEE 1680
DB 1621 KESQEKYQFLKMTAVNETQEKMCETIEHLKQBPETQKLNLENETENIRLTQILHENLEE 1680
QY 1681 MRSVTKERDRLRSVEETLKVVERDOLKENIRETITRDLEKQEBELKIVHMLKHEOETIDKL 1740
DB 1681 MRSVTKERDRLRSVEETLKVVERDOLKENIRETITRDLEKQEBELKIVHMLKHEOETIDKL 1740
QY 1741 RGIIVSEKTNIEISNNQKDLHSNDALKAQDLKIQELRIAHMLHKEQOETIDKRGIVSEK 1800
DB 1741 RGIIVSEKTNIEISNNQKDLHSNDALKAQDLKIQELRIAHMLHKEQOETIDKRGIVSEK 1800
QY 1801 TDKLSNNQKOLENSNAKLOEKIQELKANEHQLITLKQVNETQKVSMEOLKKQIKDQS 1860
DB 1801 TDKLSNNQKOLENSNAKLOEKIQELKANEHQLITLKQVNETQKVSMEOLKKQIKDQS 1860

QY 1861 LTLTKLENIENLAQELHENLEEMKSVMKERDNLRRVEETLKLRRDQLKESLQETKARDL 1920
Db 1861 LTLTKLENIENLAQELHENLEEMKSVMKERDNLRRVEETLKLRRDQLKESLQETKARDL 1920
QY 1921 EIQOELKTARMLSEKHETVDKLREKISEKTIQISDQKOLDKSKOELQKKIQBLQKEL 1980
Db 1921 EIQOELKTARMLSEKHETVDKLREKISEKTIQISDQKOLDKSKOELQKKIQBLQKEL 1980
QY 1981 QLLRVKEDVNMHKKINEMEQLKQFEPNYLCKCEMDFQTKKLHESLEIRIVAKERD 2040
Db 1981 QLLRVKEDVNMHKKINEMEQLKQFEPNYLCKCEMDFQTKKLHESLEIRIVAKERD 2040
QY 2041 ELRRIKESLWBERDQFIATLRMIARDQNHQVPEKRLSDGQOHLMESLREKCSRIKE 2100
Db 2041 ELRRIKESLWBERDQFIATLRMIARDQNHQVPEKRLSDGQOHLMESLREKCSRIKE 2100
QY 2101 LLKRYSEMDDHYECLNRLSLDLKEIEFHRIMKGLKVLVSVYTKIKESQHECINKFEMDF 2160
Db 2101 LLKRYSEMDDHYECLNRLSLDLKEIEFHRIMKGLKVLVSVYTKIKESQHECINKFEMDF 2160
QY 2161 IDEVEKQKELLIKIQLHQDCDVPESRELRLDKLNQNDLHIEILKDFSESEPSIKTEF 2220
Db 2161 IDEVEKQKELLIKIQLHQDCDVPESRELRLDKLNQNDLHIEILKDFSESEPSIKTEF 2220
QY 2221 QQVLSNRKEMTOFLEEWLNTFRDIEKLKNGIQKENDRICOVNFFNNRIIIMNESTEPE 2280
Db 2221 QQVLSNRKEMTOFLEEWLNTFRDIEKLKNGIQKENDRICOVNFFNNRIIIMNESTEPE 2280
QY 2281 ERSATISKWEQDLKSLKXNEKLFKNYQTLKTSLSAGAQVNPTQDNKNPHVTSRATOL 2340
Db 2281 ERSATISKWEQDLKSLKXNEKLFKNYQTLKTSLSAGAQVNPTQDNKNPHVTSRATOL 2340
QY 2341 TTEKIRELENSLHEAKESAMHESKLIKMOKELEVNDIIAKLOAKVHESNKCLEKTET 2400
Db 2341 TTEKIRELENSLHEAKESAMHESKLIKMOKELEVNDIIAKLOAKVHESNKCLEKTET 2400
QY 2401 IQVLQDKVALGAKPYKEEIEDLQWKLVIDLERMKNAKEPEKEISATKATVEYQKEVIRL 2460
Db 2401 IQVLQDKVALGAKPYKEEIEDLQWKLVIDLERMKNAKEPEKEISATKATVEYQKEVIRL 2460
QY 2461 LRENLRSSQAQTSVISEHTDPPQNKPLTCGGSGIVONTKALILKSEHIRELKEISK 2520
Db 2461 LRENLRSSQAQTSVISEHTDPPQNKPLTCGGSGIVONTKALILKSEHIRELKEISK 2520
QY 2521 LKQNEQLIKQKNELLSNNQHLNSNEVKTWKERTLKREAHQVTCENSPKSPKVTGTASK 2580
Db 2521 LKQNEQLIKQKNELLSNNQHLNSNEVKTWKERTLKREAHQVTCENSPKSPKVTGTASK 2580
QY 2581 KQITPSQCKERNLQDPVPKESPKCFDSDRSKSLPSPHPVRFVFNDSLSGLCPEVQVAGAE 2640
Db 2581 KQITPSQCKERNLQDPVPKESPKCFDSDRSKSLPSPHPVRFVFNDSLSGLCPEVQVAGAE 2640
QY 2641 SVDSQPGPWHASGKDVPECKTQ 2663
Db 2641 SVDSQPGPWHASGKDVPECKTQ 2663

RESULT 2
ID AAM40883
XX AAM40883 standard; protein; 2688 AA.
AC AAM40883;
XX AAM40883;
DT 22-OCT-2001 (first entry)
DE Human polypeptide SEQ ID NO 5814.
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

KW leukaemia.
XX Homo sapiens.
OS WO200153312-A1.
FN 26-JUL-2001.
PD 26-DEC-2000; 2000WO-US034263.
PF 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren P, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR N-PSDB; AAI60039.
XX Novel nucleic acids and polypeptides, useful for treating disorders such
as central nervous system injuries.
XX Example 2; SEQ ID NO 5814; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cycostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX Sequence 2688 AA;
SQ

Query Match 98.8%; Score 13170; DB 4; Length 2688;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 2643; Conservative 3; Mismatches 17; Indels 4; Gaps 4;

QY 1 MAEAGAVAVCVVRPLNRSREESLGTAQVYKTDNNVIYQVDSKSFNDFVHGNETTK 60
Db 22 MAEAGAVAVCVVRPLNRSREESLGTAQVYKTDNNVIYQVDSKSFNDFVHGNETTK 81
QY 61 NYVEIAAPIIDSAIQGYNGTIFAYGOTASGKTYTMGSEDLGVIPRA-IHDIP-QKIK 118
Db 82 NYVEIAAPIIDSAIQGYNGTIFAYGOTASGKTYTMGSEDLGVIPQGFHGFHSQKIX 141
QY 119 K-FPDREFLLRVSYMEIYNETITDLCGTQKKPLIREDVNRNRYVADLTVEVVYTSW 177
Db 142 EYFLDREFLLRVSYMEIYNETITDLCGTQKKPLIREDVNRNRYVADLTVEVVYTSW 201
QY 178 ALKWTTKGKSRHGETKQNRSSRSHITFRMILESREKGEPSNCEGSKVSHLNLVDLA 237
Db 202 ALKWTTKGKSRHGETKQNRSSRSHITFRMILESREKGEPSNCEGSKVSHLNLVDLA 261
QY 238 GSERAQGTGAAGVRLKEGCNINRSLFILGQVKKLSDGQVGGFINRYSKSLTRILQNSLG 297

||||| 262 GSERAAQTGAAGVRLKEGGINRSLFILQVQIKLSDQVGGFINYRDSKLTIRLQNSLG 321
QY 298 GNPKTRIICTITPVSFDETLTALQFASAKYKMKNTPPYNEVSTDEALLKRYRKEIMDLKK 357
Db 322 GNPKTRIICTITPVSFDETLTALQFASAKYKMKNTPPYNEVSTDEALLKRYRKEIMDLKK 381
QY 358 QLEVSLETRAQAEKQOQLAQLLEKDLLOKQVNEKIENTRMLVTSSTLQOELKAKR 417
Db 382 QLEVSLETRAQAEKQOQLAQLLEKDLLOKQVNEKIENTRMLVTSSTLQOELKAKR 441
QY 418 KRRVTWCLGKINKMKNYSADQFNIPITNTTKTHKLSINLREIDESVCSSESDVFSNTLD 477
Db 442 KRRVTWCLGKINKMKNYSADQFNIPITNTTKTHKLSINLREIDESVCSSESDVFSNTLD 501
QY 478 TLESEIENWPATKLINOENIESELNSLRADYNLDVYEQLTEKEEMELKLEKEDIDDEF 537
Db 502 TLESEIENWPATKLINOENIESELNSLRADYNLDVYEQLTEKEEMELKLEKEDIDDEF 561
QY 538 EALERKTKKQDBMOLIIHEISNLKNLVKREVYNQDLENELSSKVLLREKEDQIKKLQY 597
Db 562 EALERKTKKQDBMOLIIHEISNLKNLVKREVYNQDLENELSSKVLLREKEDQIKKLQY 621
QY 598 IDSQKLENIKMDLSYLSIESIEDPKQMKOTLFDATVVALDAKRESAFLSENLELKERKKE 657
Db 622 IDSQKLENIKMDLSYLSIESIEDPKQMKOTLFDATVVALDAKRESAFLSENLELKERKKE 681
QY 658 LATTYKQWENDIQLOSOLBAKKQWQVLEKLOSAPNEITKLTSLIDGKVPKOLLCLNE 717
Db 682 LATTYKQWENDIQLOSOLBAKKQWQVLEKLOSAPNEITKLTSLIDGKVPKOLLCLNE 741
QY 718 LEGKTTDLQKELNKEVEENEALREEVILLSELKSLPSEVERLRKEIQDKSEELHIITSEK 777
Db 742 LEGKTTDLQKELNKEVEENEALREEVILLSELKSLPSEVERLRKEIQDKSEELHIITSEK 801
QY 778 DKLFSEVVHKSERVQGLLEIFEIKTKDODLATTOSNYKSTDOEONFKTLHMDPEQKYKWL 837
Db 802 DKLFSEVVHKSERVQGLLEIFEIKTKDODLATTOSNYKSTDOEONFKTLHMDPEQKYKWL 861
QY 838 EENERNQOEIVNLSKEAQFDSLSGALKTEL SVKTOELQEKTRVQERLNEMEQLKBOLE 897
Db 862 EENERNQOEIVNLSKEAQFDSLSGALKTEL SVKTOELQEKTRVQERLNEMEQLKBOLE 921
QY 898 NRDSPLQTVVERKTLITELKQOTLEEVKTLTQEKDDLKQLOESLQIERDQKSDIHDVTN 957
Db 922 NRDSPLQTVVERKTLITELKQOTLEEVKTLTQEKDDLKQLOESLQIERDQKSDIHDVTN 981
QY 958 MNIDTQOELRNALLESUKQHOETINTLSKISSEVSRLNHEENTGETKDBFOQKMGIDK 1017
Db 982 MNIDTQOELRNALLESUKQHOETINTLSKISSEVSRLNHEENTGETKDBFOQKMGIDK 1041
QY 1018 KQDLEAKNTQTLTADVKNONEIIEQQRKIFSLIQEKNELOQMLSVTAERKQKTLDKENI 1077
Db 1042 KQDLEAKNTQTLTADVKNONEIIEQQRKIFSLIQEKNELOQMLSVTAERKQKTLDKENI 1101
QY 1078 EMTIENQOELRLIGDELKKQOEIVAQEKHAIKKEGELSRTCDRLAEVEFEKLKESQOQL 1137
Db 1102 EMTIENQOELRLIGDELKKQOEIVAQEKHAIKKEGELSRTCDRLAEVEFEKLKESQOQL 1161
QY 1138 EKOQQLNVOEEMSEMOKKINEIENIKNELKNKELTLEHMETERLELAOKLNENYEVKS 1197
Db 1162 EKOQQLNVOEEMSEMOKKINEIENIKNELKNKELTLEHMETERLELAOKLNENYEVKS 1221
QY 1198 ITKERKVLKELQSFETERDHLRGYIREIATGLQTKELKIAHILKHEOETIDELRRS 1257
Db 1222 ITKERKVLKELQSFETERDHLRGYIREIATGLQTKELKIAHILKHEOETIDELRRS 1281
QY 1258 VSEKTAQIINTQDLEKSHTKLOPEIIPVLEHEQELLPNVKVSVSETQMTNLELLTQOSTT 1317
Db 1282 VSEKTAQIINTQDLEKSHTKLOPEIIPVLEHEQELLPNVKVSVSETQMTNLELLTQOSTT 1341
QY 1318 KQSTTLARIEMERIRLNEKFOESQOEIKSLTKERDNLTKIKKALEVNDIITAKLQAKVHESNKCLEKT 1377

1342 KDSTTLARIEMERIRLNEKFOESQOEIKSLTKERDNLTKIKKALEVNDIITAKLQAKVHESNKCLEKT 1401
QY 1378 KIQESQSQSQSUNMKQKONETTKIVSEMEOFPKPKDSALLARIETIEMIGLSKLOESHEDEM 1437
Db 1402 KIQESQSQSQSUNMKQKONETTKIVSEMEOFPKPKDSALLARIETIEMIGLSKLOESHEDEM 1461
QY 1438 KSVAKEXDILORLOEVLOSQSDQKLENIKEIVAKHLETEEBELKVAHCLCKEQUEBTINELR 1497
Db 1462 KSVAKEXDILORLOEVLOSQSDQKLENIKEIVAKHLETEEBELKVAHCLCKEQUEBTINELR 1521
QY 1498 VNLSEKETEISTIQKQLEAINDKLQNKIOETIYEKEEQINIKOISEVQBNVNLQKFOKBEHR 1557
Db 1522 VNLSEKETEISTIQKQLEAINDKLQNKIOETIYEKEEQINIKOISEVQBNVNLQKFOKBEHR 1581
QY 1558 KAKOSALQSIKESKMLELTNRLQESQOEIOIIMKEKEPMKRYOEALQIERDQKENTKEIV 1617
Db 1582 KAKOSALQSIKESKMLELTNRLQESQOEIOIIMKEKEPMKRYOEALQIERDQKENTKEIV 1641
QY 1618 AKMESQOEKYOFLKMTAVNETQEKMEIEHLKQFQETQKLNLENIENTENIRLQILHEN 1677
Db 1642 AKMESQOEKYOFLKMTAVNETQEKMEIEHLKQFQETQKLNLENIENTENIRLQILHEN 1701
QY 1678 LEEMRSVTKERDDLSRVEETLKVVERDQKLENLRETITRDLKQOEELKIVHMHLEKHQETI 1737
Db 1702 LEEMRSVTKERDDLSRVEETLKVVERDQKLENLRETITRDLKQOEELKIVHMHLEKHQETI 1761
QY 1738 DKLRGIVSEKTNETISNMKQOLEHNSNDALKAQDLKIOBELRIAHMHLEKQOFTIDKLRGIV 1797
Db 1762 DKLRGIVSEKTNETISNMKQOLEHNSNDALKAQDLKIOBELRIAHMHLEKQOFTIDKLRGIV 1821
QY 1798 SEKTDKLSNMOKDLENSNAKLOEKIOELKANEHOLITLTKDVNETQKQVSEMEQLKQIK 1857
Db 1822 SEKTDKLSNMOKDLENSNAKLOEKIOELKANEHOLITLTKDVNETQKQVSEMEQLKQIK 1881
QY 1858 DQSLTSLKLEIENLNAQELHENLEEMKSVNKRDNLRVRBETLKERDQKESLOFTKA 1917
Db 1882 DQSLTSLKLEIENLNAQELHENLEEMKSVNKRDNLRVRBETLKERDQKESLOFTKA 1941
QY 1918 RDLETOQELKTYARMLSKHEKETVDKLRKISEKTTIQISDIQKDLQSKDELQKIQELOK 1977
Db 1942 RDLETOQELKTYARMLSKHEKETVDKLRKISEKTTIQISDIQKDLQSKDELQKIQELOK 2001
QY 1978 KELQLLRVKEDVNMHSHKKINEMEQKKOPEPNYLCCKEMDNFQTKLHESLEEIRIVAK 2037
Db 2002 KELQLLRVKEDVNMHSHKKINEMEQKKOPEPNYLCCKEMDNFQTKLHESLEEIRIVAK 2061
QY 2038 ERDELRRIKESLKMEDQFIATLRMIARDRQNHQVPEKRLSDGQQLHESLREKCSR 2097
Db 2062 ERDELRRIKESLKMEDQFIATLRMIARDRQNHQVPEKRLSDGQQLHESLREKCSR 2121
QY 2098 IKELLKRYSEMDDHVECLNRLSLDLKEIETEFHRIKMKLYVLSYVTIKIEEQHECINKFE 2157
Db 2122 IKELLKRYSEMDDHVECLNRLSLDLKEIETEFHRIKMKLYVLSYVTIKIEEQHECINKFE 2181
QY 2158 MFDIPEVEKQKELLIKIQLHQDCDVPSELRLDLKLNQNMDLHIEILKDFSESEFPSTK 2217
Db 2182 MFDIPEVEKQKELLIKIQLHQDCDVPSELRLDLKLNQNMDLHIEILKDFSESEFPSTK 2241
QY 2218 TEFQOVLNKRKEMTOFLEEWLNTRFDIKLNKNGIOKENDRICOVNFFNRIAINVEST 2277
Db 2242 TEFQOVLNKRKEMTOFLEEWLNTRFDIKLNKNGIOKENDRICOVNFFNRIAINVEST 2301
QY 2278 EFERSATISKEWEQDLKSLKEKNEKLPKNYQTLTKTSLASGACVNTTQDNKNPHVTSRA 2337
Db 2302 EFERSATISKEWEQDLKSLKEKNEKLPKNYQTLTKTSLASGACVNTTQDNKNPHVTSRA 2361
QY 2338 TQJTTTEKIRELENSLHEAKESAMHESKIIKMKQKELEVNDIITAKLQAKVHESNKCLEKT 2397
Db 2362 TQJTTTEKIRELENSLHEAKESAMHESKIIKMKQKELEVNDIITAKLQAKVHESNKCLEKT 2421
QY 2398 KETIQVLQDKVALGAKPYKEIEIDLKMKLVKIDLEKMKNAKEPEKISATKATVEYQKEV 2457
Db 2422 KETIQVLQDKVALGAKPYKEIEIDLKMKLVKIDLEKMKNAKEPEKISATKATVEYQKEV 2481

QY 2458 IRLLENLRSSQAQDTSVISEHTDPPSNKPLTCGGSGIVQNTKALILKSEHIRLEKE 2517
 Db |||||
 QY 2482 IRLLENLRSSQAQDTSVISEHTDPPSNKPLTCGGSGIVQNTKALILKSEHIRLEKE 2541
 Db |||||
 QY 2518 ISKLKQNEQLIKQKNEQLSNNQHLSEVETWERTLKRAHKQVTCNSPKSPKVTGTA 2577
 Db |||||
 QY 2542 ISKLKQNEQLIKQKNEQLSNNQHLSEVETWERTLKRAHKQVTCNSPKSPKVTGTA 2601
 Db |||||
 QY 2578 SKKKQITPSQCKERNLQDPVPKSPKSCFFDSRSKSLPSHPVRYFDNSSIGLCPQVONA 2637
 Db |||||
 QY 2602 SKKKQITPSQCKERNLQDPVPKSPKSCFFDSRSKSLPSHPVRYFDNSSIGLCPQVONA 2661
 Db |||||
 QY 2638 GABSVDSQPGPW-HASSGKDVPCKTQ 2663
 Db |||||
 QY 2662 GABSVDSQPGPWARLFQCKDVPCKTQ 2688
 Db |||||

RESULT 3

ABG06505
 ID ABG06505 standard; protein; 2633 AA.

AC ABG06505;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #6496.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

PN WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-PSDB; AAS70692.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

XX Claim 20; SEQ ID NO 36864; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic

CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 2633 AA;

Query Match 98.5%; Score 13131; DB 4; Length 2633;
 Best Local Similarity 99.4%; Pred. No. 0;
 Matches 2630; Conservative 0; Mismatches 2; Indels 14; Gaps 1;

QY 1 MAEAGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDGSKSFNFVRVHGNHTTK 60
 Db |||||
 QY 1 MAEAGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDGSKSFNFVRVHGNHTTK 60
 Db |||||
 QY 61 NYVEETAAPIIDSAIOGYNGTIFAYGQTASGKTYTMMGSEHDHGVIPRAIHDFQIKKF 120
 Db |||||
 QY 61 NYVEETAAPIIDSAIOGYNGTIFAYGQTASGKTYTMMGSEHDHGVIPRAIHDFQIKKF 120
 Db |||||
 QY 121 PDREFLLRVSYMEIYNETITDLCGQTKMKPLIIRDVNRYVADLTTEVVYTSEMAIK 180
 Db |||||
 QY 121 PDREFLLRVSYMEIYNETITDLCGQTKMKPLIIRDVNRYVADLTTEVVYTSEMAIK 180
 Db |||||
 QY 181 WITKGEKSRHYGETKMNQSSRSHTIFRMILESRKGEPSNCGSVKSHNLVLDLAGSE 240
 Db |||||
 QY 181 WITKGEKSRHYGETKMNQSSRSHTIFRMILESRKGEPSNCGSVKSHNLVLDLAGSE 240
 Db |||||
 QY 241 RAAQTGAAGVRLKEGNCINRSFILQVVIKKSDDGVGGFINYRDSKLTILQNSLGGNP 300
 Db |||||
 QY 241 RAAQTGAAGVRLKEGNCINRSFILQVVIKKSDDGVGGFINYRDSKLTILQNSLGGNP 300
 Db |||||
 QY 301 KTRIICTITPVSFDETLTALQFASTAKYMKNTPYNEVSTDEALLKRYKEIMDLKKOLE 360
 Db |||||
 QY 301 KTRIICTITPVSFDETLTALQFASTAKYMKNTPYNEVSTDEALLKRYKEIMDLKKOLE 360
 Db |||||
 QY 361 EYSLETRAQAMEKQDLAQLLEKDLQKQVONEKIENLRMLVTSSSLTQOELKARKKR 420
 Db |||||
 QY 361 EYSLETRAQAMEKQDLAQLLEKDLQKQVONEKIENLRMLVTSSSLTQOELKARKKR 420
 Db |||||
 QY 421 VTWCLGKINMKNSYADQFNPTNITTKHKLINLRLEIDESVCSDESNTLDTLS 480
 Db |||||
 QY 421 VTWCLGKINMKNSYADQFNPTNITTKHKLINLRLEIDESVCSDESNTLDTLS 480
 Db |||||
 QY 481 ELEMNPATKLINQENIESELNSIRADYDNLVDYQLRTEKEEMELKLEKNDLDEFAL 540
 Db |||||
 QY 481 ELEMNPATKLINQENIESELNSIRADYDNLVDYQLRTEKEEMELKLEKNDLDEFAL 540
 Db |||||
 QY 541 ERKTKKQEMQLIHEISNLKNIKLVHREVYNQDLENELSKVELLEKEQIKKIQBYIDS 600
 Db |||||
 QY 541 ERKTKKQEMQLIHEISNLKNIKLVHREVYNQDLENELSKVELLEKEQIKKIQBYIDS 600
 Db |||||
 QY 601 QKLENIKMDLSYLSIESIEDPKQKOTLFDATETVALDAKRESAFLRSNLELKEKKELAT 660
 Db |||||
 QY 601 QKLENIKMDLSYLSIESIEDPKQKOTLFDATETVALDAKRESAFLRSNLELKEKKELAT 660
 Db |||||
 QY 661 TYKOMENDIQYQSOLEAKKQVQDLEKELQSAFNEITKLSIDGKVPKDLCLNLEEG 720
 Db |||||
 QY 661 TYKOMENDIQYQSOLEAKKQVQDLEKELQSAFNEITKLSIDGKVPKDLCLNLEEG 720
 Db |||||
 QY 721 KITDLQKELNKEVEENEALREEVILLSELKSLPSVERLRKEIQKSEELHITSEKDKL 780
 Db |||||
 QY 721 KITDLQKELNKEVEENEALREEVILLSELKSLPSVERLRKEIQKSEELHITSEKDKL 780
 Db |||||
 QY 781 FSEVVHKESRVQGLLEETGKTKNDLATTQSNYKSTQDFQNFPTLHMDPEQYKVVLEEN 840
 Db |||||
 QY 841 FSEVVHKESRVQGLLEETGKTKNDLATTQSNYKSTQDFQNFPTLHMDPEQYKVVLEEN 840
 Db |||||
 QY 841 FSEVVHKESRVQGLLEETGKTKNDLATTQSNYKSTQDFQNFPTLHMDPEQYKVVLEEN 840
 Db |||||
 QY 841 FSEVVHKESRVQGLLEETGKTKNDLATTQSNYKSTQDFQNFPTLHMDPEQYKVVLEEN 840
 Db |||||
 QY 901 SPLOTVEREKLITTEKLOOTLEEVKTLTOEKDDKQLQESLQIERDQKSDHIDVNNNTI 960
 Db |||||

Db 901 SPLQTVREKTLITEKLOQTLEEVKTLTQEKDDLKQLOESQIERRDQKSDIHDTVNNMI 960
 QY 961 DTQOLRNALSKQHOETINTLTKSISEVSRNLHMEENTGETKDFQOKWGDKKQD 1020
 Db 961 DTQOLRNALSKQHOETINTLTKSISEVSRNLHMEENTGETKDFQOKWGDKKQD 1020
 QY 1021 LEAKNTQTLTADVNDNIIIEQQRKIFSLIQEKNELQOMLESVIAEKEQKLTDLKENIEMT 1080
 Db 1021 LEAKNTQTLTADVNDNIIIEQQRKIFSLIQEKNELQOMLESVIAEKEQKLTDLKENIEMT 1080
 QY 1081 IENQELRLGLDELKKQOEIVAQEKNHAIKEGELSRCTDLRAVBEKLEKSKOQLOEKQ 1140
 Db 1081 IENQELRLGLDELKKQOEIVAQEKNHAIKEGELSRCTDLRAVBEKLEKSKOQLOEKQ 1140
 QY 1141 QQLLNQVEMSEMOKKINEIENLNKELNKLTLTLEHMETRELELAQKLNENYEVKSIK 1200
 Db 1141 QQLLNQVEMSEMOKKINEIENLNKELNKLTLTLEHMETRELELAQKLNENYEVKSIK 1200
 QY 1201 ERKVLKELQSFETERDHLRGYIREIEATGLQTKBELKIAHILKEHQETIDELRRSVSE 1260
 Db 1201 ERKVLKELQSFETERDHLRGYIREIEATGLQTKBELKIAHILKEHQETIDELRRSVSE 1260
 QY 1261 KTAQIINTQDLEKSHTKLOEIPVLHBEQELLPNVKVSETQETNNELELLTEOSTTKDS 1320
 Db 1261 KTAQIINTQDLEKSHTKLOEIPVLHBEQELLPNVKVSETQETNNELELLTEOSTTKDS 1320
 QY 1321 TTLARIEMERLRLNEKFOESQEEIKSLTKERDNLKTIKEALEVKHDQKHEHIRETLAKIQ 1380
 Db 1321 TTLARIEMERLRLNEKFOESQEEIKSLTKERDNLKTIKEALEVKHDQKHEHIRETLAKIQ 1380
 QY 1381 ESQSKQESQSLNWKENDNTTKIVSEMEQFKPKDSALLRIEIMGLSKRLQESHDEMKSV 1440
 Db 1381 ESQSKQESQSLNWKENDNTTKIVSEMEQFKPKDSALLRIEIMGLSKRLQESHDEMKSV 1440
 QY 1441 AKEKDDLRQLEVLQESDQKLENIKEIVAKHLETEBELKVAHCCLEKEQETINELRVNL 1500
 Db 1441 AKEKDDLRQLEVLQESDQKLENIKEIVAKHLETEBELKVAHCCLEKEQETINELRVNL 1500
 QY 1501 SEKETEISTIOKQLEAINDKLQNKIOEIVAKHLETEBELKVAHCCLEKEQETINELRVNL 1560
 Db 1501 SEKETEISTIOKQLEAINDKLQNKIOEIVAKHLETEBELKVAHCCLEKEQETINELRVNL 1560
 QY 1561 DSALQSIERSKMLTNRNQESQEEIQIMKEKEEMKRVQOALQIERDQKENTKEIVAM 1620
 Db 1561 DSALQSIERSKMLTNRNQESQEEIQIMKEKEEMKRVQOALQIERDQKENTKEIVAM 1620
 QY 1621 KESQKEFYQFLKMTAVNETQKMCIEIHLKEQFETQKUNLENIETENLTQILHENLEE 1680
 Db 1621 KESQKEFYQFLKMTAVNETQKMCIEIHLKEQFETQKUNLENIETENLTQILHENLEE 1680
 QY 1681 MRSVTKERDDLRSVETLKVVERDQKLENIETITRDLEKQEBELKIVHMLKEHQETIDKL 1740
 Db 1681 MRSVTKERDDLRSVETLKVVERDQKLENIETITRDLEKQEBELKIVHMLKEHQETIDKL 1740
 QY 1741 RGIVSEKTEINISNMQKDLHSNDALKAQDLKTOEELRTAHMHLKEQOETIDKLRGIVSEK 1800
 Db 1741 RGIVSEKTEINISNMQKDLHSNDALKAQDLKTOEELRTAHMHLKEQOETIDKLRGIVSEK 1800
 QY 1801 TDKLSNMQKDLHSNAKLOEKIOELKANEHOLITIKKQVNETQKKVSEMEQKKQIKQDS 1860
 Db 1801 TDKLSNMQKDLHSNAKLOEKIOELKANEHOLITIKKQVNETQKKVSEMEQKKQIKQDS 1860
 QY 1861 LTLKLEIENLNLAQELHENLEEMKSVMKERNLRVEETLKLERDQKESQETKARDL 1920
 Db 1861 LTLKLEIENLNLAQELHENLEEMKSVMKERNLRVEETLKLERDQKESQETKARDL 1920
 QY 1921 ETQOELKTARMLSKHEKHTVDKLREKISEKTTIQISDIQKDLKSDQELQKQIOELQKEL 1980
 Db 1921 ETQOELKTARMLSKHEKHTVDKLREKISEKTTIQISDIQKDLKSDQELQKQIOELQKEL 1980
 QY 1981 QLLLRVKEDVNMHSHKKNEMEQKKQEPNVLCKCENDNFQTKKLHESLEETIRIVAKERD 2040
 Db 1967 QLLLRVKEDVNMHSHKKNEMEQKKQEPNVLCKCENDNFQTKKLHESLEETIRIVAKERD 2026

QY 2041 ELRRIKESLQWRDQFIATIREMIARDQRONHVKPEKRLSDGQOHLMESLREKCSRIKE 2100
 Db 2027 ELRRIKESLQWRDQFIATIREMIARDQRONHVKPEKRLSDGQOHLMESLREKCSRIKE 2086
 QY 2101 LLKRYSEMDDHYECLNLSLDLEKEIEFHRIIMKKLVLSVVTIKIEQHECINKFMDF 2160
 Db 2087 LLKRYSEMDDHYECLNLSLDLEKEIEFHRIIMKKLVLSVVTIKIEQHECINKFMDF 2146
 QY 2161 IDEVEKQKELLIKIQLHQDCDVPFSRELRLKLNQNDLHIEELKDFSESEFFPSIKTEF 2220
 Db 2147 IDEVEKQKELLIKIQLHQDCDVPFSRELRLKLNQNDLHIEELKDFSESEFFPSIKTEF 2206
 QY 2221 QQVLSNRKEMTOFLEEWLNTRFDIEKLKNGIOKENDRICQVNNFNFRILIAIMNESETEFE 2280
 Db 2207 QQVLSNRKEMTOFLEEWLNTRFDIEKLKNGIOKENDRICQVNNFNFRILIAIMNESETEFE 2266
 QY 2281 ERSATISKEWBDLKSLEKNEKLFKNVQTLKTSLASGAQVNPPTQDNKNPHVTSRAATQL 2340
 Db 2267 ERSATISKEWBDLKSLEKNEKLFKNVQTLKTSLASGAQVNPPTQDNKNPHVTSRAATQL 2326
 QY 2341 TTEKIRELENSLHSAKESAMHESKIIIMQKELEVTNDIIAKLQAKVHESNCKLEKTKET 2400
 Db 2327 TTEKIRELENSLHSAKESAMHESKIIIMQKELEVTNDIIAKLQAKVHESNCKLEKTKET 2386
 QY 2401 IQVLODKVALCAKPYKEIEIDLKMKLVKIDLEKMKNAKEPEKISATKATVEYQKEVIRL 2460
 Db 2387 IQVLODKVALCAKPYKEIEIDLKMKLVKIDLEKMKNAKEPEKISATKATVEYQKEVIRL 2446
 QY 2461 LRENLRSQQAQDTSVISEHTDPOPSNKPITCGGSGGIVQNTKALILKSEHIRELKEISK 2520
 Db 2447 LRENLRSQQAQDTSVISEHTDPOPSNKPITCGGSGGIVQNTKALILKSEHIRELKEISK 2506
 QY 2521 LKQONEQLIKQKNELLSNNHLSNEVTKWERTLKREAHKQVTCENSFKSPKVTGTASKK 2580
 Db 2507 LKQONEQLIKQKNELLSNNHLSNEVTKWERTLKREAHKQVTCENSFKSPKVTGTASKK 2566
 QY 2581 KQITPSQCKERNLQDPVPKESPKSCFFDSRSKSLPSHPVRYFDMSSILGLCEPVQNAAGAE 2640
 Db 2567 KQITPSQCKERNLQDPVPKESPKSCFFDSRSKSLPSHPVRYFDMSSILGLCEPVQNAAGAE 2626
 QY 2641 SVDSOP 2646
 Db 2627 SVDSOP 2632

RESULT 4

AA01632
 ID AA01632 standard; protein; 2954 AA.

AC AA01632;

DT 22-JUN-1999 (first entry)

DE Amino acid sequence of centromere-associated protein-E (CENP-E).

KW CENP-E; centromere-associated protein-E; ATPase activity;
 plus end-directed microtubule motor activity; chromosome congression;
 microtubule binding activity; chromosome movement; mitosis;
 cell proliferation; tumor; metastasis; vascular malfunction;
 inflammatory disease; immune disease; angiogenesis; hypertension;
 restenosis; fungal infection; selective herbicide; fungicide;
 insecticide; plant growth regulator; activator; cancer cell marker.

OS Xenopus sp.

PN WO9913061-A1.

XX 18-MAR-1999.

PF 10-SEP-1998; 98WO-US019231.

PR 11-SEP-1997; 97US-0058645P.

Db 1618 LLEELHOKNEQKLLAHEKNELEQAQVELKCEVHLMKMSIESKLSLESQHEKHDTQ 1677
Qy 1478 ELKVAHCCLEKEQEBETINELRVNLSKE-----TELSTTQKOLEANDKLNKIQEILYEKEE 1533
Db 1678 QL-----LALKQOMQVVTQEKELQOTHEHLTAEDVHLKENITE-LGLNFKNEAQKQTKTBQ 1732
Qy 1534 QL--NIKOISVQENVN-ELKOFKHEKAKDSALQSI---ESKMLELTNRLQESQEBEQI 1587
Db 1733 CLINENKELESQRLQCEIELMKSLSKQSALETLESQKVINL-----NQMEM 1785
Qy 1588 MIKEKEBMRVQEAQALQDQKENTKEIVAMKESQF--KBYQFLQMTAVNETQKMC 1645
Db 1786 VMEELKNSQRTVIAERDQDLDRESVMSIETODDLKKAQEAQOQKQVQELTSQ 1845
Qy 1646 IEHLKQEPOTKMLNENITENIRLTQILHENLEBMSVTKERDRLRSVEETLKVVERDOL 1705
Db 1846 ISVLQEKISL-----LEN-----QMLY-NVATVKTETLERDQDLSQKHLSEIETL 1891
Qy 1706 KENLRITITROLEKQELKIVHMLKHEQETID---KLRGIVSEKTEINSMQKDLHSN 1762
Db 1892 SLSUKEK-EFALQEAQKDK-----ADAARKIIDITEKISNIEEQLLOQATNLKYL-YER 1944
Qy 1763 DALKAQDLKIQEBELIAHMLKQEQETIDKLRGIVSEKTKLNSMQKDLNSAKLQEKI 1822
Db 1945 ESL-----IQCKEQLALNTEHLRETLKSKDLALGKMEQERDEAANKVIALTEKMSLEBOI 2000
Qy 1823 QE-----LXANEHOLITLKKDVNETQKVSMEOLKQIKDOQLTSLK-----EIE 1869
Db 2001 NENVTLKEGEGEKETFYLRPSKQOSSQMEELRESLTKDQLQLEBAKEISEATNEIK 2060
Qy 1870 NL-----NLAQELHENLEMSVMKERNLNRVEETLKLRLDQKESLQETKARDLEIQ 1924
Db 2061 NLTKISSLEBEILQNASILNEANVERENLRHSKQOLVSELEQLSILT---KSRDHAFQA 2117
Qy 1925 ELKTARMLSKHEKTVDKLRKISKTIQISDIQKDLK---SKBELQKQIQLQ----- 1976
Db 2118 -----SKREK---DEAVNKIASIABEIKILTKEMDFRDSKESLQEQSHLSSEELCT 2166
Qy 1977 -KKEQLLLR-VKEDVNMS-HKKINEMEQKKQFEPNYLCKCEMDNFOLTKLHESLEIR 2033
Db 2167 YKTELQMLKQKQEDINNLAKVKEVDEL-----LOHLSLKEQLDQIQ 2210
Qy 2034 IVAKERDELRLTKESLQME-RDQFIATREMIARDRONHQVKEPKLLSDGOQHLMESLR 2092
Db 2211 MELR-NEKLRNYELCEKMDIMEKEISVLELM-----QNEPQOE-EDDVAERMDILE 2259
Qy 2093 EKCSRIKELLKRYSEM-DHYECLNRLSLDLKEIEFHR----- 2130
Db 2260 SRNQEIQELMEKISAVYSRQHTLLSLSELQKETEHAHKMINTIKESLSTLSRSGSL 2319
Qy 2131 -----IMKKLAVLSYVTKIEEQHECINKPEMDFIDBEKOKELLIKIQLH 2177
Db 2320 QTEHVKLNTQLOTLLNKFKV--YRTAAVKEHSLIKDYEKDLAAQKHDELRLQQL 2377
Qy 2178 QOD-----CDVPSRELRLKLNQMDL-----HIEBILKDFSEF--FPSIKTEFOQVILS 2225
Db 2378 EQHGRKWSDSASBELKFCIEFLNELLFKKANIIOQVQDDFSEVQVFLNQVQSTLQEELE 2437
Qy 2226 NRKEMTQFLEEMINTRFDEKLNKGQKENDRICQVNNFNRRILAIMNESTEFERSAT 2285
Db 2438 HKKGFMQWLEEFGLHDVDAKLSSEGQENRRITASTIQLTKRLKAWQSKI---QREIT 2494
Qy 2286 I-SKEWEQDLKSLKEKNEKLFKNYOTLTKTSLAGQVN-----PTQONKNPHVTSRA 2337
Db 2495 VYINQPEAKLQEKQKQNKELMRMEHHGPGASVMEENARLLGILKTVQDE-----SKX 2548
Qy 2338 TQLTTEKIRELSLHSAKESAMHESKSIIMQKQELVTNDIATL---QAKVHESNKCL 2394
Db 2549 LQ---SRIKMLENELNLVKDDAMHKGKVAILOQL--LSRNAEALNMQVLTQKQDNL 2604
Qy 2395 EKYKETIOVLQDKVALGAKYKEIEDLKKMLKIDLEKMKNAKEFEKESATKATVEYQ 2454

Db 2605 QAAMKEIENLQKVAKGAVPYKEEIDNLTKVVKIEMEKIKYSKATDQETAYLKSCLDK 2664
Qy 2455 KEVIRLLREMLRSQQADTSVISEHTDDPOPSNKPPLTCGGSGTIVONTKALILKSEHRL 2514
Db 2665 EELRLRLKEBLRRAQADNDITVCVPDYOQKASTFPVTCGGSGTIVQSTAMLVLOSEKAL 2724
Qy 2515 EKEISKLKQONEQL-----JKONELLSNNHLSNEVKTWKERTLKREAHKQV 2562
Db 2725 ERELSHYKKYVHLSRMTSSSEDRKTKAKSDAHSHTGSHRSPHKTYR---HGPV 2781
Qy 2563 TCE-----NSPK-----SPKVTGT-----ASK 2579
Db 2782 TPERSEMPSLHLGSPKSESSTKRVSPNRSEIYSLVMSPGKTMGHKILSPKVGHLK 2841
Qy 2580 KKQITP-----SCKERNLQDPVPKESPKSCFF 2607
Db 2842 KRALSPNRSEMPTOHVISPGKTLGKHLNLTSTFLDNLSPPCKQKQVQENL--NSPKGLKLF 2899
Qy 2608 DSRKSLSPSPHPVYFQNSISGLCPQVQNAQESVDSPGPWHAASSGKDVPECKT 2662
Db 2900 DVKSKMPY-CPSQOFFNSKGLGDFSELNTAESNDKSAENWWYBAKETAPECKT 2953

RESULT 5
ABB62322
ID ABB62322 standard; protein; 2013 AA.
XX
AC ABB62322;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 13758.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
PN
PD 27-SEP-2001.
XX
PP 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR N-PSDB; ABL06425.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signaling and cell-cell
interactions.
XX
PS Disclosure; SEQ ID NO 13758; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signaling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
sequences (AB101840-AB116175) and the encoded proteins (AB57737-
AB577072). The sequence data for this patent did not form part of the
printed specification, but was obtained in electronic format directly
from WIPO at fcp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2013 AA;

Db 1835 -----KVDEVTCEKLFDMQSEVQNSKVQEL----- 1864

QY 1991 MSHKKINEMELKQFEPNYLCKEMDNFOLTKKLESLEIRIVAKERDELRIKESL- 2049

Db 1865 -----ISCEBELSTL-----KKEASFQSEK-----ESMD--RISSILLEKRNLEKLC 1908

QY 2050 -----KMERDQFIATIREMIARDQ-----NHQVKEKRL--SPGQOHLMESL 2091

Db 1909 SANDIVAKLEVE--IAALRPKSLDRNPVPRKSTFSEIRKRRISVHDERRQSYNDV 1966

QY 2092 RE 2093

Db 1967 RE 1968

RESULT 6

ABB61012

ID ABB61012 standard; protein; 1931 AA.

XX ABB61012;

AC ABB61012;

XX 26-MAR-2002 (first entry)

DT Drosophila melanogaster polypeptide SEQ ID NO 9828.

DE Drosophila; developmental biology; cell signalling; insecticide;

XX Drosophila; developmental biology; cell signalling; insecticide;

XX Drosophila melanogaster.

OS WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

PA Venter JC, Adams M, Li PWD, Myers EW;

PI WPI; 2001-656860/75.

DR N-PSDB; ABL05115.

DR New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions.

PS Disclosure; SEQ ID NO 9828; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutic and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

CC ABB72072). The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1931 AA;

QY Query Match 9.6%; Score 1286; DB 4; Length 1931;

Db Best Local Similarity 23.8%; Pred. No. 3.2e-47;

Matches 527; Conservative 418; Mismatches 767; Indels 500; Gaps 84;

QY 6 AVACVVRPLNSREESLGTAQVYKTDNNVIYQVDS--KSFNEDRVFPHGNETTKNVY 63

Db 8 SIQVCIKVRPCFGLTSL-----WQVKEGRSIQLADSHAEPYVDYVDFEGASNQEVF 60

QY 64 BEAAPIIDSAIOGYNCTIFAYGOTASGKTYTMMGSEDLGLVPIRAIHDIPOKIKKPPDR 123

Db 61 DRMAKHIVHACMONGSNTIIFAYGOTSSGKYITMMGDEONPGVMVLAKEIFQOLSSETER 120

QY 124 EFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVVYADLTETEVVYTSMALKWIT 183

Db 121 DFLLRVGYIHYNEKIYDYL--NKNQDLKHESGNGIVNV--NCKESIVTSEDDLLRQLY 177

QY 184 KGEKSRHYGTQMMQORSRSHHTIPRMILESEKPEPCSGSVKSHNLNVLVLASRAA 243

Db 178 MGNERVVGETNMNERSRSHAIPIRIIESKSDHSDN--DTVKQSVLSVLVLASGEQVD 235

QY 244 QTGAAGVRLKEGCNINESLFILOVHKLSGQVGGFINVRDSDKLTIRLQNSLGNPKTR 303

Db 236 PADHAS-----SLMIFRNLVKSLSVDSKPNFSDSKLPRIMLPSLGGNVLTS 284

QY 304 IICITIPVSFDETLTALQFASTAKYMKNTPTVNEVSTDEALLKRYKEIMDLKKQLEVS 363

Db 285 IICITIPSFVESSSTISFGTCAKKIRCKPQCKIDSETTM--RQLD--- 330

QY 364 LETRAQAMEKDOLAQLLEBKLOLVONEKTENITRMLVTSSSLTLOQELKAKRRVTW 423

Db 331 ---RGISMLKDKLAK-----KKIKNE-----SOLVL-QELEGRIKRDML- 365

QY 424 CLGKINKMNSNYADQFNIPNTITTKHLINLREIDESVCSGESDVFSNTDLSRIE 483

Db 366 -----KIVSSASLDLRLQKRRTWLTAS-----GSEGDAPVALPEPESR 408

QY 484 WNPATKILNQENIESELSNLRADYDNLVLYQEUTKEBEMELKUKENNDLDEFALERK 543

Db 409 LPRPSKLTNLPKPLFQRRGIAPKAGICKTLKEKRLQTDNMD----- 450

QY 544 TKDOEMOLIHEISNLKMLVHREVYNODLN-----ELSSKV-----ELLREKEDQ 590

Db 451 TMPGRAKQLGRETSRIEPSVMMSKKYQESVFNCDAPQTEISALTASNOVAKETIEKYE 510

QY 591 IKKQLEVIDSOKLENIKMDLSVLESIEDPKMQKTLFDAETVALDRAKRESLENLE 650

Db 511 VARLKETIERLEWENGA-----VN 530

QY 651 LKERMKELATTYKOMENDIQLYQSOLEAKKQVQVLEKELQSAFNEITKLTSLIDGKVPK 710

Db 531 LGEQFETHKAKSKQMEEL--LSSISEKOSTIVSLQOSLE-----ELSR 572

QY 711 DLLCNLELEGKITDLOKELNKEVER--NEALREEVIL--LSELKSLPSEVERLRKEIQ 764

Db 573 DVLNRSK-EDQMRSMCPLESSCERICNKCLELERLLPLASASGLDSVACQDFQLRSEIA 631

QY 765 DKSEELHIITSEKDKLFSEVVHKEKRVQGLLEIGTKTDLATTQSNYKSTQDFQNFK- 823

Db 632 ATRMKLESMLSTFSSHASCEVSQKTTDCRKLSEIQTSTAHDDFQQLQEKYNNLKHKWSQKL 691

QY 824 ---TLHMDP---EQYKVMYLENE---RMNQEI NVLSKEAKQFSSGALKTSLKYKIQ 873

Db 692 AIDTWQVDVNTIQQYQLQODEYRHLERLSDQCOLOQDENSCKLOAEIGTLKE----RVE 747

QY 874 ELQEKTRV---QERLNEHQLEQENRDSPLQ----TVREKTLITIEKLOQTLEVEKT 926

Db 748 EIHSELLEVPNDTPPEMELQNKQELKSLQKLEWFEIQLNYECLSELMSITQECDA 807

QY 927 LTQE-----KDDLQLOESLQIERDLKSDIHDVTN--MNIDTQEQRLNALSLEKQ- 975

Db 808 LREEHKQRTTNSDLESMSKSS-----GVGTCSDPENELDTLLQQTKLKSKIQIILT 861

QY 976 -----HOETINTLKS-KISEVSVRNLMHEENTGETKDEFFQKQVMGLDKKQDLE 1022

Db 862 DYSGGRRLLFIYHAEQDQSVPSLKLCLPEAKYL---EGDGKQHDASDSVFL-----KGLFK 914

QY 1023 AKNTQTLTADVKDNEIIEQORK---IFSLIOEKNELOQMLESVIAEKOQLTKLKENIEM 1079

Db 915 CORFOVTKINEQNLVKEEDMRDIIIFOLKQEVGDKKNLIE-----BEKEVINNLRAQITS 970

QY 1080 -----TIENQCE-ELRLIGDELKKQEIIVA---QEKHNAIKKEGELSRTCDRLAEVEPKLK 1130

Db 971 LNOETIKNOAKTKILCELOTKDTQVQANKQSOEVLTKTSLAHLKSKVCELOKLE 1030
QY 1131 EKS--QOQOEKQOQOLLNVQSEMSEMOKINEIENLK-----NELKNKLTLEHMET-- 1179
Db 1031 KQSEDEKISELOSDIGEISECCLSMELKLDIIVNQABELRPDUQOESGVELQHSHTTA 1090
QY 1180 -EKELEAOKLNNYEVKSIKERVKLKELQKSFETERDHLR-----GYI 1223
Db 1091 EESLNVKPIQEQOTERTLTTEYERRI--EQLEESLQRAQAEELSILEKXKTDENKSLQLEYM 1149
QY 1224 RETEATGLQKEELKAHILHKEHOETIDELRSVSEKTAQIINTQDLEKSHTKLOEIP 1283
Db 1150 AKTISENENRSPRAYCLDKTKRYEQLOQTNEKLASVT-----TQCOVHLD 1200
QY 1284 VLHEQELLPNVKVSQETQMTNELELLTQSTTKDSTTLARTEMERLR--LNEKQESQ 1341
Db 1201 VIKRSLQ-----EKITQAEKRNEL-----AVHKAELEKIRETLKEKSSYK 1243
QY 1342 BEIKSLTKERDNLKTIKEALEVKHDQKEHIRETLAKIQESQSQBOQSLNMKEKONETTK 1401
Db 1244 EKLQAEERD--KEISR-LEV-----MRNTIAELHKTNSDREVEL-----EGVK 1285
QY 1402 IVSEMEQFKPK---DSALLRIEIMGLSKRLQESHDMSKSAVAKK--DDLQRL----- 1450
Db 1286 ----MEKCOLKLYDKSML--ELEQ-LQCTAD-QKSSDLLPGSSNENIDDLQKCDQYQD 1338
QY 1451 QEVQESDQLKENIKEIVAKH-----LETEBELKVAHCKLKEQETIN 1494
Db 1339 LELLRGEKAEALLSEIQKINGQHSNTIKLEETAEEMITITTKLEB--RCEIAKLETPK 1396
QY 1495 ELRVNLSEKETEISTTKOLEA-----INDKLQNKIOEIVEKREQ 1534
Db 1397 SKEADIKK---ALHCAQLRLHAYDKLVCEYERLKGCLSDSNKLSNLEKQVERLHAQLA 1453
QY 1535 L-----NKKQI--SEVQENVNELKQFKEHKAQDSALQIESQMLBLTNRLQK----- 1580
Db 1454 LQEGISGRDSEITKQRLSELKDAIDENKTVREAKVGLNSLKAVQENMSAQEGQFKQKAD 1513
QY 1581 ---SQQEIQIMIKEEMKRVQEQALQIERDQKENTKEIVAKMKSQOEKEYQFLKMTAVN 1637
Db 1514 IKGSDVDELQIKLSQEV-----RDHLESNEELKRLKQAEQLQNMVMDKERKLN 1563
QY 1638 ETQEMKCEIEHLKEQF---ETQKINLETETENIRLTQI--LHENLEEMSVTKERDILRS 1693
Db 1564 SS-----LRQDFDKLEQTKLDLE---EQLRAKKEIDRRSKELGEVTKDCENIRS 1610
QY 1694 -----VEETLKVERDOLK---ENLRRTITDLEKQELKIVHMLKHEQETIDK 1739
Db 1611 DLEAQTDFLKERETINUTISDLRLHNEQLLETSTKNYLSDDITAAANNLEMKNLHDLTK 1670
QY 1740 ----LRGIVSEKTNISNMQKLEHSDALKQAQDLKIODELRIAHMLKEQOETIDKLRG 1795
Db 1671 ECKSLRSDIQKEEYFQTKQLLDDETINLKEENRAMEEKLSSGNKALKEDE---KLRS 1727
QY 1796 IVSEKTKLSNMQKLENSNAKLBK-----IQELKANEHQTLTKK-----DVNE 1841
Db 1728 TLESKELIQOQKLEERLTVINEKGNKALLDAQLKSNETAFSLKAWIKQSALATEA 1787
QY 1842 TOKKVSMEQ--LKKQIQDOSLTSLKLEIENLQAQELHENLEEMKSVMKERDNLRR---- 1896
Db 1788 ANKRSLEMEQKVDKRTREYELRSTLTREINFRSEKERMDOGTISLLEDKRNLEKLC 1847
QY 1897 VEETL--KLERDQKESLQETKARDLEIQOELKTAARMLSKEHKEITVDKLEKISEKTIQIS 1955
Db 1848 VTELLAKLKEIPLAHTOKVNGDVSI--ELNSNG-----SPTPAAPV 1889
QY 1956 DIQKDLQ--KSKDELOKTKQELQKLEQLLRVKEDVNMHSHKINEMEQKQF 2006
Db 1890 ATKKPLDCNSAECPKSSSLETAE-----RQNRRTAVDENRQF 1930

RESULT 7

AAU83007
ID AAU83007 standard; protein; 240 AA.
XX AAU83007;
XX
DT 23-APR-2002 (first entry)
XX
XX Human homologue of YP1 protein target for antifungal compound.
XX
XX Antifungal; fungal gene transcription; RPC34; POP3; TFA2; NAB2; MPT1;
KW MTR2; BOS1; POL30; RSA2; SQT1; MTW1; TPB1; SPC98; BFR2; RNAL; GCD7; SKI6;
KW NIP1; LCP5; NCE103; ECO1; ORC2; CNS1; YPDI; TIM10; SRB4; yeast; fungus.
XX
OS Homo sapiens.
XX
XX WO200202055-A2.
PN
XX
XX 10-JAN-2002.
XX
XX 28-JUN-2001; 2001WO-US020592.
PF
XX
XX 29-JUN-2000; 2000US-0215164P.
PR
XX 10-AUG-2000; 2000US-0224457P.
XX
XX (ANAD-) ANADYS PHARM INC.
XX
XX Moore J, Buurman ET, Desilva T, Harris S, Komarnitsky S;
PI Mendillo M, Moore D, McCoy M, Sanderson K, Haq T, Zhu S, Long F;
PI Davidov E, Thompson CM;
XX
DR WPI; 2002-147962/19.
N-PSDB; ABK32895.
XX
XX
PT Screening candidate antifungal compound for interaction with essential
PT protein, modulation of essential protein activity, binding to essential
PT protein, by contacting protein with test compound and determining
PT effects.
XX
XX Claim 1; Fig 79; 522pp; English.
XX
XX The invention describes a method of screening a candidate antifungal
CC compound for interaction with essential proteins (EP) or for modulation
CC of EP activity e.g fungal gene transcription. The proteins tested in the
CC invention include RPC34, POP3, TFA2, NAB2, MPT1, MTR2, BOS1, POL30, RSA2,
CC SQT1, MTW1, TPB1, SPC98, BFR2, RNAL, SKI6, NIP1, LCP5, NCE103,
CC ECO1, ORC2, CNS1, YPDI, TIM10 and SRB4 from S. cerevisiae, C. albicans
CC and human homologues. The method involves contacting a culture with one
CC or more test compounds and determining the effects on the growth or
CC viability of the culture of cells which preferably comprises fungal cells
CC or yeast cells. Preferably the identified compounds interact with, or
CC modulate (preferably inhibit) activity of C. albicans EP. The inhibitor
CC compounds identified by the method are useful for preventing or
CC inhibiting fungal, particularly C. albicans growth in culture or in a
CC mammal. The antifungal agents interact with essential fungal elements
CC that can be used to treat fungal infection by preventing the growth and
CC preferentially killing the fungi, but does not inhibit the biological
CC activity of mammalian homologues. This amino acid sequence represents a
CC target protein used to test the antifungal compounds, described in the
CC method of the invention
XX
SQ Sequence 240 AA;

Query Match 8.9%; Score 1183; DB 5; Length 240;
Best Local Similarity 100.0%; Pred. No. 9.1e-44;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1801 TDKLSNMQKLENSNAKIQEIKELKANEHQTLTKKDVNETQKVSMEQIKQIKQDS 1860
Db 1 TDKLSNMQKLENSNAKIQEIKELKANEHQTLTKKDVNETQKVSMEQIKQIKQDS 60
QY 1861 LTLSKLEIENLQAQELHENLEEMKSVMKERDNLRRVEETLKLRDQKESLQETKARDL 1920
Db 61 LTLSKLEIENLQAQELHENLEEMKSVMKERDNLRRVEETLKLRDQKESLQETKARDL 120

QY 1921 EIOBELKTARMLSKHEKTVDKLREKISEKTIQISDIQKDLDSKDELOKKIOBLOKEL 1980
 Db 121 EIOBELKTARMLSKHEKTVDKLREKISEKTIQISDIQKDLDSKDELOKKIOBLOKEL 180
 QY 1981 QLLRVKEDVNMHSHKKNEMQLKKQKQFPNLYCKCEMDNFOLTKKLHESLEIRIVAKERD 2040
 Db 181 QLLRVKEDVNMHSHKKNEMQLKKQKQFPNLYCKCEMDNFOLTKKLHESLEIRIVAKERD 240

RESULT 8
 ADE56037
 ID ADE56037 standard; protein; 3259 AA.
 XX
 AC ADE56037;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Human Protein CAA53052, SEQ ID NO 1876.
 XX
 KW Human; pain; neuronal tissue; gene therapy;
 KW spinal segmental nerve injury; chronic constriction injury; CCI;
 KW spared nerve injury; SNI; Chung.
 XX
 OS Homo sapiens.
 XX
 FN WO2003016475-A2.
 XX
 PD 27-FEB-2003.
 XX
 PF 14-AUG-2002; 2002WO-US025765.
 XX
 FR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX
 XX (GEMO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 XX
 PA Woolf C, D'urso D, Befort K, Costigan M;
 PI WPI; 2003-268312/26.
 DR GENBANK; CAA53052.
 XX

PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX
 PS Claim 1; Page; 1017pp; English.
 XX
 CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound comprising the one or more
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a human protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 3259 AA;
 Query Match 8.5%; Score 1127.5; DB 7; Length 3259;
 Best Local Similarity 20.9%; Pred. No. 4e-40;
 Matches 643; Conservative 596; Mismatches 1011; Indels 831; Gaps 143;

QY 87 QTASGKTYTMMGSEHGLVPIRAIHDFIKIKFPDPREFLLRVSYMIYETITDLCGT 146
 Db 447 ETASQTSFPDYNVNEGTAQVTEENIASQKRVLENEKGAALLSSIE-----LEELKAEN 501
 QY 147 QKMKPLIIRBDVNRNVVADLTVEVVTSEMAKWKTKGKSRHYGETKMKQSRSH-T 205
 Db 502 EKLSSQITLLEAQRNRTGEADREVSEISIVDIANKRSSAEES---GDVLENTFSQKHKE 558
 QY 206 IFRMILESREKGPSPNCEGSKVSHLNL-----VDLAGSRAAQGAAGVR 251
 Db 559 LSVILLEKMAQE-----ETAFKLQOGKRAEADHEVLDKEMKQMEGEGIAPIK 610
 QY 252 LKEGCNINRSLFILGQVIKKLSD--GVGGGFINTYRDSKLTRILQNSLGNPKTRIICTTP 310
 Db 611 MK-----VF-----LEDTGQDFPLMPNEESSLPAAVEKQEASTEHQSR---TSEE 651
 QY 311 VSPDETTLAQFASTA-----KYMKNTPYNEVSTDEALLKRYRKEIMDLK-----KOLE 360
 Db 652 ISLND--AGVELKSTKQDGKSLSAVPIGQCHQDE--LERLKSQILLELNFHKAQIY 707
 QY 361 EVSLETRAQAMEKDOLLAQLLEE---KDLLQKVQNEKIENLRLMTSSSLTLOQELKAK 416
 Db 708 EKNLDEKAK--EISNLNLQIEEFKKNADNNSAFTALSEERDQLLSQVKELSMVTELRAQ 765
 QY 417 RKRRTVTCGLKINKMNSYADQFNIPNTIITTKTKLSINLRLREIDRESCVESDFVSNLT 476
 Db 766 VK-QLEVMNLAEARQRRLDYESTQAHNDNLLTEQIHSLSIA-----KSKDVKIEVLQNEL 819
 QY 477 DTLSEIENPATKLLNQ-----ENISEL-----NSLRADYDNLVLDVDEQLRTEKE 522
 Db 820 DDV-QLOFSEQSTLIRSLQSLQNKQSEVLEGAERVRHISKVEELSQALSQLEIYTKM 878
 QY 523 EMELKLEKNDLDEF-EALERK-----TKK-----DOEMQLTHEISNLK---NL 562
 Db 879 D-QLLEKRRDVTETLQTTIEEKDQVTEISFSMTEKVMQVLENEKESLSGLVEIKTLKEQNL 937
 QY 563 VKHREVYNQDL---ENELSSKVE-----LUREKEDQTKKLOE 596
 Db 938 LSRABEAKKQVEEDNEVSSGLKQNYDEMSPAQISKEELQHEFDLLKKNEQKRKLQA 997
 QY 597 YIDSQK--LENIKMDLSVLSIESIEDPKQMKQTLFPAETVALDAKRESAPLRSENLELKEK 654
 Db 998 ALINRKEQLQVSR-LEELANLKDSEKKEIPLSETERGEVEEDK-----ENKEYSEK 1049
 QY 655 MKELATTYKQWENDIQLYQSQLEAKKMQQ---VDLEKEL--QSAFNEITK-LTSLJDGKV 708
 Db 1050 ----CVTSKQBEIYVLQTTIEKEVELOHIRKDLKLAEEQFQALVKQMNQTLQDKT 1105
 QY 709 PKDLNLCNLELEGKITDLQKELNKEVEENE---AUREEVILLS-----E 748
 Db 1106 NQIDLLQAEISENQAIQKLTISNTDASDGSVALVKETVWISPPCTGSSEHMKPELEEK 1165
 QY 749 LKSLSFVERLRKETEODKSEELHITSEKDKLFSEVHKESRVQGLLEIGTKDGLAAT 808
 Db 1166 ILALEKEBQLOKLOE-----AUTSRK-AILKKAQKE---RHUREELKQKDDYNRL 1215
 QY 809 QSNYKSTDOEFONFKTLHMDFFQKVMVLEENRMQEIWNLSKEAQKDFSSIGALKTEL 868
 Db 1216 QSQFDBQSENENIG-----DQLRQLQIVRESIDGKLKLP--STDQOESCSPTGLEEPL 1267
 QY 869 SYNTQELQKTRVQERL-----NEMEQLKEQLENRDSPLQTVREK 910
 Db 1268 -FKATE-QHHTQPVLESNLCPDWPSHSEDASALQGGTSAQIKAKLKE-----IEAEK 1318

[illegible]

840	QO	VAEVAQKQVCTELDAHKIQVODLMQOLEKQNSEMEQKVKSLTVQVYESKLEDGCKQE	899
1279	QY	QEEIPVLHBEQELLPNVKKVSETQETWNELELLTEQ-STTKOSTTILARIEMERLRNLNEKF	1337
900	QY	EQYKQILVEKENM---ILQMREGQK--KEIFILTKLSAKEDSIHILNEEYETKFNQ---	952
1338	QY	QESQETIKSLTKERDNLKTTIKALEVXKHDLQKXHIRETILAKIQESQKQOSLNMMKEDN	1397
953	QY	EKKMEKVKQAKEMQ--ETLKKQLDQEAKLKKELENTALEL-----SQEKQFNAK----	1002
1398	QY	ETTKIVSEMEQFPK--DSALLRIEIMLGLSKRLOESHDE-----MKSVAKEKD	1446
1003	QY	-----MLEMAQANSAGISDAVSLETNQKEQIESLTVHRELNVDVISIWEKLNQQAEE	1057
1447	QY	LQRLQEVLOSQDOLKENIKEIVAKHLETEEEELKVAHCLKE-----QEETINELRVNLSE	1502
1058	QY	LQEHETLOQEKEGEVAELAKQILLFGCEKEENNKEITWLKERGVQODTTNLNEQOLQK	1117
1503	QY	KETEISTI-----QKOLEAINDKLQNKIQEIIYEKEEQLNIKQISEVOENVNELKQFKEHRKA	1559
1118	QY	KSAHVSILAQDETCLKAHLKLEVDLANK-----SLKENTFLOEQVLVELKMLAEDDKR	1169
1560	QY	KDSALQIESKMLELTNRLQESQEEIOIMIKE-----EMKRVQEAQIQRBDQ	1608
1170	QY	-----ELTSKLKTTIDEEFOSLKSHEKSNKSLDEKSLFKLSSELAIQOLDI	1219
1609	QY	LKENTKEIV-AKMK-----SQEKEYQFL-----KMTAVNET-QEKMCEIHLKEQFE	1654
1220	QY	CKKTEALLEAKTNELINISSKTNAILSRISHCQHRTTKVGEALLIKTCTVSELAQLR	1279
1655	QY	---TKMLNLEIETENIRLTQILHNLBEMRSVTKERDDLRSVBETLKV-----RDQKENLR	1710
1280	QY	QLTEEQNTLNSFO--QATHQLEBKENQIKSMKADIESLVTKEKALQKEGNGOQAAASEK	1337
1711	QY	ETIYRDLEKQBELKIVMHMLKEHOETIDKLGRIVSEKTNESNMOKDLEHSNDALKAQDL	1770
1338	QY	ESCITQLKKE-----USENINAVTLMKEELKEKVELLSSLSQJ-----TDL	1379
1771	QY	KIQEELRTAHMHLKEQOETIDKLGRIVSEKTDKLSNMKOLSENSNAKLOEKIOLKANEH	1830
1380	QY	NVQLQNSIS-----LSEKEAAISLRLKQYDEEKCILLDQVDL-----SF	1419
1831	QY	QLITLKQDVNETQKVSMEQJ-----KXQIKDQ-----SLTUSKLEIENLNAQE	1876
1420	QY	KVDTLKSE-----KISALEQVDDWSNKFSEWKKKAQSRFTQHONTVKELQIQLELSKE	1473
1877	QY	LHENLBEMKSVKMERDNLRVBEETILKLERDQKLSQETKARDLEIOELTATMLMSKEH	1936
1474	QY	AYEKOEDNLILKBELDQONKAPDLCLUGEMEDDK-SMKKESNLTELEKLSQTARIM----	1528
1937	QY	KETVTKLREKISEKTIQ-----SDIOKDLDKSD-----BLQKKIQBELQK-KELQLLRVKED	1988
1529	QY	-----ELEDHITQKTIEIESLNEVLKNYNQOKDIEHKELVQKLQHFQELGEBEKDNVKE-	158
1989	QY	VNMSHKKINEMBOLKKQFEPNYLCKCEMDNFOLTKKHSHESLEETIRIVAKER-DELRRIKE	204
1583	QY	---AEEKILUTL-----NQVTSMAELE-----TKK--KELEHVNLSVKSKEEELKALEL	1627
2048	QY	SLKMERDQFIATLREMIARDRONHQVKEPKRLSDGQOHLMESLEKCSRICKELLKRYSE	210
1628	QY	RLEESAAKLAELKR-----KAPQKTAATKKQ-LUSOMEK-----EEQYKKGTE	167
2108	QY	MODHYECLNRLSIDL-EKEIEPHRMKMLKVVLS-----YVTVKIBEQ	214
1672	QY	--SH---LSELNLTQOERERFVHILEELKLSVSESSQSETLIVPRSAKNVAAYTEQBEADS	172
2150	QY	HECINKFDMDFIDEV-----EKOK-----ELLKIQ-----	217
1727	QY	QGCQVQKTYEETKISVLQRNLTEKELQLRVGOBEKEETVSSHFMRCQYQBELIKLEHAENAK	178
2176	QY	-----HLQDCDVPREL-----RDLKONMDLHIEETILKDPSESE	221
1787	QY	QHEQDSQMIGHLOEBELBENKKNYSLIVAQHVKEGGKNNIQAOKNLENVEDDVQKTLQEXE	184

QY	1108	ATKSGELS--RT-----CDRLAEVBEKLEKESQQLQEQQLLNQOEEMSEMOKKINEIE	1161	QY	2108	MDDHYECLNRLSLD--EKEIEFHRIIMKKLYVLS-----VYTKIKREQ	2149
Db	738	SIIEKEVSIQRTKALQDQINQINQLELLKERDKHLEHQA-----HVE	780	Db	1672	--SH---LSELNTKLOEREREVHILEKLSVSSQSETLIIVPSAKNVAAYTEQFEADS	1726
QY	1162	NLKNELKNKELTLEHMETERLELAQKL--NENTVEVKSITKERRVKELQKSFETERDHLR	1220	QY	2150	HECINKFEMDFIDEV-----EKQK-----ELLKIQ-----	2175
Db	781	NLEADIKRSEBELQ--QASAKLDVFOYSQATHEQTKAYEQLAQQLKLDLITERILLT	839	Db	1727	QGCYQVTEKISVLRQNLTEKEKILQVRGOBEKTEVSSHFMRCQYQERLIKLEHAEAK	1786
QY	1221	GYIRELEATGLQTKBELKTAHILKHQETIDELRSVSEKTRAIINTQD--LEKSHTKL	1278	QY	2176	-----HIQDCDVPSPREL-----RDKLQNMDLHIEELKDFSESE	2212
Db	840	KQVAEVAQKQVCTELDAHKIQVQDLMQOLEKQNSEMEQKVSALTQVVEKSLDGNKEQ	899	Db	1787	QHEQDSMIGHLOBELEENKKNYSLIQAQVHEKGGKNNQAQONLENVDPDVQKTLQKE	1846
QY	1279	QEEIPVHBEQELLPNVKVSETOETWNELELLETEQ--STTKDSTTLARIEMERLRINEKF	1337	QY	2213	F-----PSIKTEFOVLNRRKEMTOFLEEWLNTRFDIEKLKNGIOKENDRICOVNFPN	2266
Db	900	EQTQKQILVEXENM--ILOVREGOK--KEIEITQKLSAKEDSIHILNEEYEFKNQ--	952	Db	1847	LTCQIILEQKIK--ELDSCLVRQKEVHVEEELTSKY--EKQ--ALQOMDGR-----	1893
QY	1338	QESQERISLTKERDNLTKTKEALEVHQDLKEHIRETLAKIQESQSQSQSINMKKON	1397	QY	2267	NRIIIMNTEPEERSATISKWEQDLKSLKNEKLFKNYQTLKTSLSAGQAVNPTTQ	2326
Db	953	EKKMEKVQKAKEMQ--ETLKKLLDQEAKLKKELENTALEL-----SQEKQFNAK--	1002	Db	1894	NKPTLEENTEESKSHLVQPKLLSNMEA--QHNDLEFK-----LAGAREKQKL	1942
QY	1398	ETTKIVSEMQFKPK--DSALLRIEIMGLSKRLQESHDE-----MKSVAKEKDD	1446	QY	2327	DNKNPHVTSRATQITTEKIRELENSLHEAKESAMHESKIIIMQKELEVTND-----IIA	2381
Db	1003	-----MLEMAQANSAGISDAVSRLTQKQIESLTVHRRRLNDVTSIWEKLNQQAEE	1057	Db	1943	GKEIVRLQKDLRMLRKEHQOELEIL-----KKEYDOEREKIKQOEDELELKHSTLKQIMR	1999
QY	1447	LQRLQVLOQESDQLENIKIEIVAKHLETEELKVAHCCKE-----QETINELRNLSE	1502	QY	2382	KLOAKVHESNKCLEKT--KETIQVLQDKVALGAKYKEIEIDLKMLKVIDLEKMKNAKEF	2440
Db	1058	LQEIHEIQLOQKEQEAVALKQKILLFGCEKEMMKEITLWKEGVQDITLNEIQEQLQK	1117	Db	2000	EFNTQLOAKQOELEMTIKETINKAQEVAELSHQEBETNQLLKIAEKDDDLKRTAKRY	2059
QY	1503	KETIRISTI--OKOLEALNDKLNQIKQIYEKEBQNLNKOISVQENNVNELKQKEHKA	1559	QY	2441	EKEISATKATVYQKQVIRLLRENLRSSQQAQDTSVISEHTDPOPSNKPPLTCGGSGGIVQ	2500
Db	1118	KSAMVNSLAQDETQKHAHLEKLEVDLNL-----SLKENTFLOEQVQLVLMKLAEDKR	1169	Db	2060	BEILDARE--EEMTAKVRDIQTQLEBELQKYQOKLFOEE--NPGNDNVT-----	2104
QY	1560	KDSALQSTIESKMLELNTNLRQSQEQIOMIKEKE-----EMKVEQEAQIERDQ	1608	QY	2501	NKALILKSEHRLKEISKLKQNEQLIKQKNELSN-----QHLNNEVKTWKERTLKR	2556
Db	1170	KVS-----ELTSKLTUTDEBFQSLKSHESKNSKSLDEKSLFKKLEELAIQDII	1219	Db	2105	-----IMELQTO-----LAQKTLISDSKLKEQEFREQHINLEDRLLKY	2143
QY	1609	LKENTKEIV--AKMKE-----SQEKYQFL-----KMTAVNET--QEKMCETIEHLKEQFE	1654	QY	2557	EAHQVTCENSP	2568
Db	1220	CKKTEALLEAKTNELINISSKTNAILSRISHCHQRTYKVEALLIKTCTVSELEAQLR	1279	Db	2144	ENVAIVTGTGP	2155
QY	1655	--TOKNLNENIETENIRLTQILHENLEMRSVTKERDDLRSVETTLKVE--RDQKENLR	1710	RESULT 13			
Db	1280	QUTTEQNTLNISFO--QATHOLEENQKJSMKADIESLVTKEALQKEGNGQQAASEK	1337	ABR61601			
QY	1711	ETITRDLKQEBELKIVHMLKHEQETIDKLRGIVSEKTEIENMOKLEHSDNALKQAOL	1770	ID	ABR61601	standard; protein; 2250 AA.	
Db	1338	ESCITQLKKE-----LSEINAVTLMKBEKKEKVEISSLSKQL-----TDL	1379	XX	ABR61601;		
QY	1771	KIQEELRIAHMLKBEQETIDKLRGIVSEKTKLSNMOKDLENSNAKLOEKIQELKANEH	1830	AC	15-JAN-2004	(first entry)	
Db	1380	NVQLQNSIS-----LSEKAAISLRLKQYDEKCELDQVQDL--SF	1419	DT	Human golgin-245 splice variant 3 polypeptide.		
QY	1831	QLITLTKDVNETQKKVSEMEQL-----KKQIKQDQ-----SLTILSKLEIENLNAQE	1876	DE	Golgin-245; neurodegenerative disease; neuroprotective; nootropic;		
Db	1420	KVDTLSKE-----KISALEQVDDWSNKFSEWKKKAQSRFTQHTQNTVVKELQIQLELKSKE	1473	KW	antiparkinsonian; anticonvulsant; gene therapy; human; variant.		
QY	1877	LHENLEEMKSVMERONLRVETTLKLRDQLKESQETKARDLEIQOELKTARMLSKEH	1936	XX	Homo sapiens.		
Db	1474	AVEKDEQINLLKEELDQONKRFCLGEMEDDK--SKMEKESNLETLEKSGQTARIM--	1528	FT	Key	Location/Qualifiers	
QY	1937	KETVDKLRKISKTQI-----SDIQKDLDSKD-----ELQKTIQELQK--KELQLRLKVED	1988	FT	Misc-difference	2244	/note= "encoded by TTT"
Db	1529	-----ELEDHTIKTIEESINEVLKNYNQKQDIEHKLQKLFQHQELGEBKDNRYKE--	1582	FT	Misc-difference	2245	/note= "encoded by ACT"
QY	1989	VNMSHKKINEMEQKQKQFEPNYLCKEMDNFQTLKULHESLEIRIVAKER--DELRIKE	2047	FT	Misc-difference	2246	/note= "encoded by TCA"
Db	1583	-----ABEKILILE-----NQVYSMAELE-----TKK--KELEHVNLSVSKSEELKALD	1627	FT	Misc-difference	2247	/note= "encoded by CCT"
QY	2048	SLKMRDQFATLRMTARDQNVKPEKLLSDGOQHLMESUREKCSRIKELLKRYSE	2107	FT	Misc-difference	2248	/note= "encoded by CGC"
Db	1628	RLESESAKLAEALR-----KABQKIAAKKQ--LUSQMEEK-----EEQYKKGTE	1671	FT	Misc-difference	2249	/note= "encoded by AGT"
QY				FT	Misc-difference	2250	/note= "encoded by GGT"
Db				XX			
PN				XX			

Db	1341	EKEALQEGNQQAASEKESCITQLKKE-----LSENINAVTLMKEELKEKVEI	1391
Qy	1752	SNMQDLEHSDALKAQDLKIQBELRIAHMHLKEQOETDKLRCIYSEKTDKLSNNQKDL	1811
Db	1392	SSLSKQL-----TDLNVOLQNGSIS-----LSEKAAISSLRKQY	1425
Qy	1812	ENSNAKLOEKIQLKANEHQLIITLKKDVNETQKKVSEMEQL-----KKQIKDQ--	1859
Db	1426	DEEKCELLDQVQDL-----SFKVDITLSKE-----KISALEQVDDWSNKFSEMKKKAQSRFT	1476
Qy	1860	--SITLSKLEIENLNLAOELHLENEMKSVYKBERDNLRRVEETLKLERDQLKESLOETKA	1917
Db	1477	QHONTVKELQIQLKLSKEAYEKDQNLNKEELDQCKRFDCLKGEMEDDK-SKMEKKE	1535
Qy	1918	RDLEIQOELKTARMLSEKHEKTDVKLREKISEKTIQI-----SDIQKDLKSKD-----ELQK	1970
Db	1536	SNLETLSKQTARIM-----ELRDHITQKTIETIESNEVLKNTYQNDQIEHKELVQ	1586
Qy	1971	KIQELQK-KELQLLRVKEDVNMHKKINEMOLKKQFEPNVLCKCENDFNQITKKLHESL	2029
Db	1587	KLOHFQELGEKDNVRKE-----AEKILTLB-----NQVYSMAKALE-----TKK--KEL	1630
Qy	2030	EETRIVAKER-DELRRIKESLKMEDQFIATLRMIARDRONHVQKPEKRLLSGQOQHLM	2088
Db	1631	EHNVLVSKSEELKALEDRLESSEAAKLAEKR-----KAEQKIAALIKQ--LL	1678
Qy	2089	ESLREKSRIKELLKRYSEMDDHYECLNRLSIDL-EKEIEPHRIWKKLYVLS-----	2140
Db	1679	SOMBEK-----BEQYKKGTE--SH-----LSELNKLQEREREVHIILEEKLKSVESQSETLI	1729
Qy	2141	-----VVTIKEBQEHCIKFMDFIDEV-----EKQK-----	2168
Db	1730	VPSAKNVAATEGEADSQCQVQTYEEKISVLQNLNTEKEKLLQVRQGEKEVTSSH	1789
Qy	2169	-----ELLKIQ-----HLQDCDVPSEL-----RDLKL	2193
Db	1790	EMRCQYQERLLKLEHAAKQHDQSMIGHLOBELEKNKYSLIQAQHVKEGGKNNIOA	1849
Qy	2194	NONMDLHLEILKQSFSEF-----PSIKTFQOVLNRKEMTQFLEEWLNRFDIEKL	2247
Db	1850	KQNLNVDFDVQKTLQEKELTCQILEQKIK-ELDSCLVRQKEVHVHVEELTSKY--EKL	1906
Qy	2248	KNGIQKENDRICOVNNFFNRIIAIWNSTEPSERSATISKWEQDLKSLKNEKLPKN	2307
Db	1907	Q-ALQOMDGR-----NKTELEENTEEKSKSHLVQPKLLSNMEA--QHNDLEFK-	1953
Qy	2308	YOTLKTSLASGAQVNPPTQDNKNPHVTSRATQLTTEKTIRELNSLHLEAKESAMHYESKII	2367
Db	1954	-----LAGAEREKQKLGKEIVRLQKDLRLMRKEHQOELEIL---KKEYDQEREKIK	2002
Qy	2368	KMQKLEVTND-----IAKLOAKVHESNKCLEKT-KETIQVLQDKVALGAKPYKEEID	2421
Db	2003	QEQEDLELKHNSLTKQLMREFNTQLAQKQOELEMTIKETINKAQEVEALELHSHQETNQ	2062
Qy	2422	LKMKLVKIDLEKMKNAKEPEKETSATKATVYQKEVIRILRENLRSSQAQDTSVISEHT	2481
Db	2063	LLKKAETKDDDLKRTAKRYEELDARE---EEMTAKVRDLQTOLELQKYQOQLFOEB-	2118
Qy	2482	DFQPSNKPITCGGSGGIVONTKALILKSHIRLEKISKLQONEQIILKQKELLNSN--	2539
Db	2119	--NPGNDNVT-----IMELQTO-----LAQKTTLISDSKL	2146
Qy	2540	--QHLNVEKVTWKERTLKRAHKQVTCENSP	2568
Db	2147	KEQEFREQIHNLDRLLKKEKNVYATTVGTP	2177
RESULT 14			
ID	ABR61602	standard; protein; 2252 AA.	
XX	XX	ABR61602	
AC	ABR61602;		

XX	DT	15-JAN-2004	(first entry)
XX	XX	Human golgin-245 splice variant 4 polypeptide.	
DE	XX	Golgin-245; neurodegenerative disease; neuroprotective; neurotropic; antiparkinsonian; anticonvulsant; gene therapy; human; variant.	
KW	XX	Homo sapiens.	
OS	XX	WO2003087403-A2.	
XX	PN	23-OCT-2003.	
XX	PD	16-APR-2003; 2003WO-EP003958.	
XX	PF	16-APR-2002; 2002EP-00008553.	
XX	PR	16-APR-2002; 2002US-0372424P.	
XX	XX	(EVOT-) EVOTEC NEUROSCIENCES GMBH.	
PA	XX	Von Der Kammer H, Pohlner J;	
XX	PI	WPI; 2003-845345/78.	
XX	XX	N-PSDB; ACF58153.	
DR	XX	Diagnosing or prognosticating a neurodegenerative disease comprises comparing a level or activity of a transcription or translation product of golgin-245 gene in a test sample with a reference value of a known disease or health status.	
PT	XX	Claim 18; Fig 11; Opp; English.	
PT	XX	The invention relates to diagnosing or prognosticating a neurodegenerative disease in a subject, or determining whether a subject is at increased risk of developing the disease. The method involves determining a level and/or an activity of a transcription or translation product of a gene coding for golgin-245 and/or a fragment, derivative or variant in a sample from the subject and comparing the level and/or the activity to a reference value. The composition and methods are useful in diagnosing, preventing and/or treating neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease, Huntington's disease or amyotrophic lateral sclerosis. The recombinant, non-human animal is useful for screening, testing and validating compounds, agents and modulators in the development of diagnostics and therapeutics to treat neurodegenerative diseases, in particular Alzheimer's disease. Specific antibodies are used for detecting the pathological state of a cell in a sample from a subject. The present sequence represents a human golgin-245 splice variant 4 polypeptide	
XX	XX	Sequence 2252 AA;	
SQ	XX	Query Match 8.2%; Score 1089; DB 7; Length 2252;	
		Best Local Similarity 22.4%; Pred. No. 1.2e-38;	
		Matches 559; Conservative 503; Mismatches 849; Indels 580; Gaps 113;	
QY	344	LLKRYRKEIMDLKQLLEVSLETRAQAMEKQLAQLLEKDLQKQVNEKINLTMLVT	403
DB	1	MFKKLKQKISEQQQLQALAPAQASSNSTP-----IFMRSR	38
QY	404	SSSLTIQ-QLKAKAKRRVTVCLGINKMKNSYADQFNIPNTTKT--HKLSI-----	455
DB	39	TISFTQLDEGTPNENAST---HASKSPDSVNGSEPSIPQSGDTQSPAQLQLRVPVS	94
QY	456	-NLLRE-IDESV---CSESDFNSVNTDLTLEIWNPAT-----K	489
DB	95	ESLFRSPIKESLFRSSSKESLVRTSSRESLNRLDLSSTASFPDPMDSSEADLVGNSD	154
QY	490	LINQENI-----ESELNSLRADYNLVDYQLRTEKEEMELKKEKNLDLFEALER	542
DB	155	SLNKEQLIQLRLRMERSLSYRGKYSIELVTAYQMLQREKKLQGLLSQSD---KSIRR	210
QY	543	KTKKQEQMQLIHEISLNKLVKREYVYNQDLENELSKVELLEKEDQIKKLQEIYD--S	600

QY 2540 --OHLSEVKTWKERTLKRAHKQVTCNSP 2568
Db 2147 KEQEFREQHNLLEDRLLKYEKNVYATVGGP 2177

RESULT 15

ADE56031
ID ADE56031 standard; protein; 3187 AA.

AC ADE56031;

XX 29-JAN-2004 (first entry)

XX Rat Protein BAA05026, SEQ ID NO 1870.

XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
XX chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

XX Rattus norvegicus.

XX WO2003016475-A2.

XX 27-FEB-2003.

XX 14-AUG-2002; 2002WO-US025765.

XX 14-AUG-2001; 2001US-0312147P.

XX 01-NOV-2001; 2001US-0346382P.

XX 26-NOV-2001; 2001US-0333347P.

XX (GHEO) GEN HOSPITAL CORP.

XX (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-268312/26.
XX GENBANK; BAA05026.

XX New composition comprising two or more isolated polypeptides, useful for
XX preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat
XX or human polynucleotides or a polynucleotide which represents a fragment,
XX derivative or allelic variation of the nucleic acid sequence. Also
XX claimed are a vector comprising the novel polynucleotide, a host cell
XX comprising the vector, a method for identifying a nucleotide sequence
XX which is differentially regulated in an animal subjected to pain and a
XX kit to perform the method, an array, a method for identifying an agent
XX that increases or decreases the expression of the polynucleotide sequence
XX that is differentially expressed in neuronal tissue of a first animal
XX subjected to pain, a method for identifying a compound which regulates
XX the expression of a polynucleotide sequence which is differentially
XX expressed in an animal subjected to pain, a method for identifying a
XX compound that regulates the activity of one or more of the
XX polynucleotides, a method for producing a pharmaceutical composition, a
XX method for identifying a compound or small molecule that regulates the
XX activity in an animal of one or more of the polypeptides given in the
XX specification, a method for identifying a compound useful in treating
XX pain and a pharmaceutical composition comprising the one or more
XX polypeptides or their antibodies. The polynucleotide or the compound that
XX modulates its activity is useful for preparing a medicament for treating
XX pain (e.g. spinal segmental nerve injury (SNI)), chronic constriction
XX injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
XX therapy). The sequence presented is a rat protein (shown in Table 2 of
XX the specification) which is differentially expressed during pain. Note:
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic form directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3187 AA;

XX SQ

Query Match 8.2%; Score 1089; DB 7; Length 3187;
Best Local Similarity 20.9%; Pred. No. 1.8e-38;
Matches 634; Conservative 558; Mismatches 1012; Indels 828; Gaps 124;
QY 113 IFOKIKKFPDPREFLLRVSYMEIYNETIT-----DLLCGTQMKKPLIIREDVNRVYVA 165
Db 221 VMOKLEBEHEEALLGRAQVVDLLQKELTSAEQRNQL---SQQLLLEAEHSLRNTMEA 277
QY 166 DLTEEVVYVYSEMALKWITGK-----SRHVG-----ETKMNORS 201
Db 278 ERQESKILMEKVELEMAERKEELYQLQGLERAGQAQAELEMOYGTTLQOHHETEMEEK-- 335
QY 202 RSHTIFRMILESEKGEPSNCEGSKVSHLVNLDLAGSRAAQTGAAGVRLKEGCNINRS 261
Db 336 ---TACISLQKNEQELQSACD-ALKEENSKLIQ-EQEQQAQSAQALQLEDE----- 384
QY 262 LFLIGQVIKLSQGVGFNYRD-SKLTILQNSLGNPKRIICITIPVSPFETLTAL 320
Db 385 ---LQKSKELISQ-----FVNKNLEKHETSSQTSL-----PDVYNEGVQAV 423
QY 321 QFASTAKYMKNTPVYNEVSTDEALLKYRKEIMDLKKLEBVS-----LETRAQAMEKDQ 375
Db 424 MEESVASLQKR---VLELENEKGALLSSLELELRANEKLCRSITLLEAQNAGEADG 480
QY 376 LA-----QLLEE--KDLLQKVQNEKIENILTMLVTSSSLT-----LQOEL 413
Db 481 MVCEVSTAGIALNRSDSSTESQGVLENTFSQKHKELSVLLVEMKEAQEELAFKLSQ 540
QY 414 KAKRKR-----RVTVCLGKINMKMKNS-----NYADQF 440
Db 541 QKRPEDGYEVLDRKEVQMESEGLPSVTARDVLC---APRDKNVPAVEGEQAGMRDQH 597
QY 441 NI-----PTNIT-----TTHKL-----SINLLREI 461
Db 598 GTLEAGPLNDTGMELNSPQDPQVSKLSAPHVCLHQGELERLKTQVLESLTAKET 657
QY 462 DESVCSDESDFNTLDTLSIEIEMNPATKLLNQ--ENIESELNSLRADYDNLV-----LD- 513
Db 658 HEKNLSEK---AKEISSLTQL-----TKFKESAEEARSTLTAVCEERDQLLYRVKELDV 709
QY 514 YBQLATEKEEMELKLEKNDLDEFBALERTKKQDQEMQLIHEISNLKLVKREYVNDQL 573
Db 710 LGELRAQVRELETSLAE--AEKQRGLDYESORAHNLLTEQIHSLSLSEAKSKDVKIETL 766
QY 574 ENELSKVELLEKEDQIKKQFYDLSQKLENIKMDLSYLSIESIDPKMKQTLFPAETV 633
Db 767 QRELDGVQLQFSEQGTQIKSLQSLQTKR-----SEVLGAERMK----- 806
QY 634 ALDAKRESAFLRSENLKLEKMKELATTYKQMNENDI-QLYQSQLEAKKQXQVDLEKELQS 692
Db 807 -----DISKEMEELSQAQSKLELAQMDQLLEKKQDVET-LQQTIOE 849
QY 693 AFNEITKLTSLIDGKVPKOLLNLELGKITDLQKELNKEV-----BENEALREVILLSE 748
Db 850 KQQVIT-----ELFSMTKRWQVLENEKFSGLGVEIKTLKEQLSULSR 891
QY 749 L-----KSLPSEVERLRKGIQKSEELHITSE 776
Db 892 AEGAKREQVEDSGAESSPKHGPHESSAEPVCKEALQOELWELURKESEQRKQLQALIS 951
QY 777 KDKLFSEVHHKESRVQGLLEEIGTKTDDLATTSQNYKSTIDQEPQNFKTLHDMFEQKY--- 833
Db 952 RKELLOKVKLEELAKVREE--STKDSL--RESEKRELEEDSKN-----KDDPEKYGTS 1002
QY 834 -----KMWLEENRMNQEIIVNISKEAQKDFSSLGALKTSLSYKTOELOEKTRV-- 882
Db 1003 EWRELEVSRLTITSEKEVELEGIRDLKERAEEELQALVQRM---TQDLQNTKQKLDL 1059
QY 883 -QRLNEM-----BOLKEQLENRSDPSLQTV 906
Db 1060 LQBEITENQATIQKFTITGTMADGDSAVKETSVSPPRAGGEGHWPKELEGK---IVDL 1116

QY 907 EREKTLITEKLOTELEVYU---TQKODLKOQESLOIERDQLKSDIHDVTNNIDTQ 963
 Db 1117 EKEKTOLOKQLOEALISRAKAILKKAQKE---KHLKEELKEQDAYR-HLQBPQDQSKEN 1173
 QY 964 EQLRNALLESKQHOETINTILSKISEV---SRNLHME-----ENTGETK 1005
 Db 1174 ENIRAPILQAKESTDQQLPGTQOQEPHSGEGSLGTEGPASESDLHAAQPSHPGETA 1233
 QY 1006 D-----BFOQMMVGID-KKQDLEAKNTQTITADV--KONEIIEOQRKFSLIQKNE 1054
 Db 1234 TLQATVSVAQIQDQKLEIVEKEELELKISST-TSELTAKSEEVILLQEQINEGLEION 1292
 QY 1055 LQOMLESVIAEKQKTDKLENMTIENOEERLLGDELKQOEIVAQEKNAHAKKEGE 1114
 Db 1293 LKAASHEAKAHTLQKQEL-ESSQLKIADLEHLKTLQPELETIKHVQKQEEVSYLVGQ 1351
 QY 1115 LSRTCRLAEVSEKLEKESQ-----OLQEQKQOQLLNQOESMOKKINEIENL 1163
 Db 1352 LGEKEQTLTVQTEMEGEERLIIKALHTQLEMOAKEHEERLKQVVEICELKQPKOLE-- 1409
 QY 1164 KNEKKNELTLEHMETERLELAQKLNENYEVKSIITKERKVLKELQKSP-----E 1213
 Db 1410 -BESKAKQOLQKQALISRKALKENKSLQOLQSARDVAEHLTKSLADVESQVSVQN 1468
 QY 1214 TERDHLRG-----YIRIEATGLQTK-----BELKIAHILKHEQETIDELRR 1256
 Db 1469 QEKDALLGKALLQOERDKLIVEMDKSLLENQSLGSCESLKLALGLTEDKEKL--MKE 1526
 QY 1257 SVSEKTAQIINTQDLKSHTKLOEETPVL-----HEBQELLPN 1294
 Db 1527 LESVRSKIAESTEWQEKHELOKEVEVLQSYENVSNEABRIQHVVESVROEKQEVYAK 1586
 QY 1295 V-----KKVSETQFTMNELE-----1309
 Db 1587 LRSABSDKREKQKQDQAEQEMEEMKMKRKAQKQKQKILEEENDRLRAPAQVGA 1646
 QY 1310 -----LLTEQSTTKDSTTLARIEMRLINRKNKFOESQEBEISLTKERDNLKTIKEALRV 1363
 Db 1647 NBSMEALLSSNAKLEE--LERITLEYKTLSEFEALMAEKNTLSEETNLKLOVEAQEL 1704
 QY 1364 KH-----DOLKEHIRE-TLAKIOESQSQEQSLNKEKDNETTKIVSEMEQFKPKD 1413
 Db 1705 QOASLETTEKSDPKDVIETVAVVGSQEQDSLENAKLEDAETLLANSA---KP-- 1759
 QY 1414 SALLRIEIMLGLSRLOESHDMSVAKEDDLO-RLQEV-LQESD-----OLKENIKE 1467
 Db 1760 -----GVSETF-SSHDINNLYLOQLQKGRITAELEMEKQKORELSQTTLEKN 1807
 QY 1468 IVAKHLETEE-ELKVAHCLKEQEBETINELRVNLSEKETEISTIQKOLEAINDKLQNKIQ 1526
 Db 1808 ALLTOISAKDSKL-----LEBEVAKINMLNQIOBELSRVTKLKTAEERKDDLEERL- 1862
 QY 1527 BIYEKEQOLNKIOISEQENYNELQFKEHRKAKDSALQSTESKMLELTNRLQESQETQ 1586
 Db 1863 -----MNQIAELNGSINYQYQDVTDAQKN---EQLESEMQLNKRVCSELEERQ 1909
 QY 1587 IMIKE-----KEEMKRVQEA-----LOIERDOLKENTKEIVAKMKEQSEKQYQFLK 1632
 Db 1910 QLVKEKTVESEIRKEYNEKQGAQKPGSKIHAKELQELKKEQOEKQVQKQKOCIRVIG 1969
 QY 1633 MTAVNETQEKMB-----IHLKBE-----QFETQKLNLENTENIR 1669
 Db 1970 RISALEKTVKALEFVHTESQKDLQATKGNLAQAVEHHKKAQAEELSSFKILLDDTQSEAR 2029
 QY 1670 LTOIILHENLEEMRSVTKERDDILRSVEETLKVERDOLKENLRETIYTRDLEKQEEIKVIMH 1729
 Db 2030 ---VLADNL-----KLKELQSNKESIK---SQIKQK-DEDLRLLEQAE-----KH 2070
 QY 1730 LKEHQETIDKLRIGVSEKTEINSNQKDLHSDNALKAQDLKIOBELRIAHMLKQOET 1789
 Db 2071 RKEKKNQEKLDALHREKA-----HVEDTL-----AEIQVSLTRKDKMKELQOS 2115
 QY 1790 IDLGRGIVSEXTDKLSNMQKOLEN--SNAK-----LOEKIQELKANEHQILTLKCO 1838

Db 2116 LDSTIAQAAFTKSMSSLODDRVIDEAKWQORFGDAIQTKEEVRLKEENCNTALK-- 2173
 QY 1839 VNQTKVSEMEQKQKQIKQDQSLTSLKLEIENLNAQELHEN--LEBMKVMKERNLRR 1896
 Db 2174 -DQLRQMTIHEELK-----ITVSRLEHD-----KEIWESKAQTELQHOQKAYDKLOE 2220
 QY 1897 VEETIKLRLQKESLOETKARDLEIQOELKTARMLSKHEKHEVTDKLE-----KI 1947
 Db 2221 ENKELASQLEAAGQLYHDSKNELTKLESELKSLKQDQSTDLKNSLEKCREHNNLEGIKQ 2280
 QY 1948 SEXTIQ-----ISDIOKLDKSD--ELQKKQLOELKQKLOLLRVKE-----DVMN 1991
 Db 2281 QEADIQNKFCNCQLETDLTASRELTTRRDEHINVKEQKIIISLSGFEAEIOVAIAELHQ 2340
 QY 1992 SH-KKINEMEQKKQFEPNYLKCENMFOLTKKLHESLEIRIVAKER-----DELR 2043
 Db 2341 QHSKEIKELNLLSQEENL-TLEENKRAVEKTLQLTALTEATIKKESLEQKALQDSFV 2399
 QY 2044 RIKESLMEKDQFIATLRMIARDRONHVQPEKRLSD--GOHLMESLREKCSRIKEL 2101
 Db 2400 KSMSSLODDRDRIVSDYQOL-----EERHLSVILEKDELIDQAAEANNKKEE 2447
 QY 2102 LKRYSEMDDHYECLNELSLDLEKE-IEPHRMKKL-----KVLSVYTKIKEQHE 2151
 Db 2448 IR---GLRGHMDLNSENAKLDAELIOYRDLNEVITIKDSQORQLLEAQLOONKELRNE 2504
 QY 2152 CINKFEMDFIDEVEKQKELLIKIQLHQDCDVPSERLDLKLNONMDLHTEELKDPSES 2211
 Db 2505 CV-KLEGRLKSEAEKQSLQMSLDLQOENQGLSEIKSF-K-EQLTALHEEGALAVY-HA 2561
 QY 2212 EFPSTKTEQOVL-----SNRKEMTQFLEWLNTRFDIEKKNKGIQKENDRIQCVNNFFNN 2267
 Db 2562 QURVREVEEQKLTAAALSSQKRTVDLQEBLVCQKEASKVSEIEDKLR--ELKHLHN 2619
 QY 2268 RIIAIMNESTETEERSATISK---EWEQDLKSLKKEKNEKLFKNYQTLKTSLAGAQNPT 2324
 Db 2620 AGI-MRNETETAERVAELARDLVEMEQKLLTVTKENKOLTQAIQAFGKSMSS----- 2671
 QY 2325 TDONKNPHVTSRATQITTEKIRELENSLHEAKESAMHKESKIKMKOKELEVINDIIAKLQ 2384
 Db 2672 LQDSRD-HATELSDL-----KKYDASLAKELAKGRQDUGRESVDLSQA 2717
 QY 2385 AKVHESNCKLETKETIQVLQDKVAKGAPYKEEIEDLMMKLVKIDLEKMKNAKEFEKE- 2443
 Db 2718 PFL-----TTSSENI-----SSRLEKLNQQLISKDEQLHLSSLESSH 2755
 QY 2444 ---ISATKATVEYQKEVIRLLR--ENLRSQQAQDTSVISEHTDPQPNKPLTCGGSGI 2498
 Db 2756 NQVQSFTKAMTSLQNERDHLNNELEKFKSEBEGORSAA-----PSAASSPAE----- 2803
 QY 2499 VONTKALI--LKSEHIRLEKISKLQOQNEOLIKOKNEL-----LSNNOHLSENVKTKW 2550
 Db 2804 VOSLKAMSSLOQDRDLKELKNLOQYLOMNOEITELRPLKQALQESQDQTKALQMK 2863
 QY 2551 ER-----TLKREAHKQVTCNSPKPKVTGTASKKQIITPSQCKERNLODPV 2597
 Db 2864 BELQENLSWOHELDQLRVEKNKSWELHERRMKEQVYLMASDKDQQLGHLNLLREL----- 2919
 QY 2598 PKESPCKSCFFDSRSKSLPSPHPVYFDNSSLG 2629
 Db 2920 -----RSSOTQILPTQYQRASSG 2939

Search completed: July 29, 2004, 09:35:09
 Job time : 180.292 secs

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; Z15005; CAA78727.1; --
DR PIR; S28261; S28261.
DR HSP; P17119; 3KAR.
DR Genew; HGNC:1856; CENPE.
DR GK; Q02224; --
DR MIM; 117143; --
DR GO; GO:0005699; C:kinetochore; TAS.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0008350; F:kinetochore motor activity; TAS.
DR GO; GO:0000657; P:DNA replication and chromosome cycle; TAS.
DR GO; GO:0007079; P:mitotic chromosome movement; TAS.
DR GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
KW Motor protein; Cell division; Aff-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere; Lipoprotein; Prenylation.
FT DOMAIN 1 335 KINESIN-MOTOR.
ET DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP BIND 86 93 ATP (BY SIMILARITY).
FT LIPID 2660 2660 S-farnesyl cysteine.
SQ SEQUENCE 2663 AA; 312087 MW; CFC13880C8C8C8 CRC64;

Query Match 100.0%; Score 13329; DB 1; Length 2663;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2663; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEGAVAVCVVRPLNSREESLGETAQYVWKTNNVYVDGSGSFNDRVFGHNETTK 60
DB 1 MAEGAVAVCVVRPLNSREESLGETAQYVWKTNNVYVDGSGSFNDRVFGHNETTK 60
QY 61 NYVEIAAPIIDSAIQNGTIFAYGQTASGKTYTMGSEDLGVIPRAIHDFOKIKKF 120
DB 61 NYVEIAAPIIDSAIQNGTIFAYGQTASGKTYTMGSEDLGVIPRAIHDFOKIKKF 120
QY 121 PDRELLRVSYMEIYNETITDLCCTQKMKPLIREDVNRNVYVADLTVEEVVYTSSEMALK 180
DB 121 PDRELLRVSYMEIYNETITDLCCTQKMKPLIREDVNRNVYVADLTVEEVVYTSSEMALK 180
QY 181 WITGKSRHYGETKMNQSRSHITFRMILSRKSGPSPNCEGSKVSHLMLVDLAGSE 240
DB 181 WITGKSRHYGETKMNQSRSHITFRMILSRKSGPSPNCEGSKVSHLMLVDLAGSE 240
QY 241 RAAQTGAAGVRLKECINRSIFILGOVTKLSDGVGFINVRSKLTRIQLNSLGNP 300
DB 241 RAAQTGAAGVRLKECINRSIFILGOVTKLSDGVGFINVRSKLTRIQLNSLGNP 300
QY 301 KTRIICTTIPVSFDETLALQAPASTAKYMNTPYNEVSTDEALLKRYKEIMDLKKQLE 360
DB 301 KTRIICTTIPVSFDETLALQAPASTAKYMNTPYNEVSTDEALLKRYKEIMDLKKQLE 360
QY 361 EYSLTRQAAMEKQDLAQLEEKDLQKVONEKIENLTMVLTSSSLTLQOELKAKRKR 420
DB 361 EYSLTRQAAMEKQDLAQLEEKDLQKVONEKIENLTMVLTSSSLTLQOELKAKRKR 420
QY 421 VTWCLGKINMKNYSADQFNIPNTITTKTKLSINLREIDESVCSDFVSNLTDLIS 480
DB 421 VTWCLGKINMKNYSADQFNIPNTITTKTKLSINLREIDESVCSDFVSNLTDLIS 480
QY 481 EIEWNPATKLNQENIESELNSIRADYNLDVYEQLRTEKSEMELKAKRNDLDFEAL 540
DB 481 EIEWNPATKLNQENIESELNSIRADYNLDVYEQLRTEKSEMELKAKRNDLDFEAL 540
QY 541 ERKTKDQEMQLIHEISNLKLVKREVYNQDLENELSSKVELLREKEDQIKLQEVIDS 600

DB 541 ERKTKDQEMQLIHEISNLKLVKREVYNQDLENELSSKVELLREKEDQIKLQEVIDS 600
QY 601 QKLENIKMDLSYLSIESIEDPKMQKTLFDAETVALDAKRESAFIRSENLEKMKELAT 660
DB 601 QKLENIKMDLSYLSIESIEDPKMQKTLFDAETVALDAKRESAFIRSENLEKMKELAT 660
QY 661 TYQMENDIQLYQSLEAKKMQVDLEKELQSAFNEITKLTSLIDGKVPKDLLCNLEBEG 720
DB 661 TYQMENDIQLYQSLEAKKMQVDLEKELQSAFNEITKLTSLIDGKVPKDLLCNLEBEG 720
QY 721 KITDLOKELNKEVEENALREEVILLSELKSLPSEVERLRKEIQDKSEELHIITSEKDL 780
DB 721 KITDLOKELNKEVEENALREEVILLSELKSLPSEVERLRKEIQDKSEELHIITSEKDL 780
QY 781 FSEVVKRESRVQGLLEIGIKTKDLDLATTQSNYKSTDOERFQNFKTLHMDFOKYKMWLEEN 840
DB 781 FSEVVKRESRVQGLLEIGIKTKDLDLATTQSNYKSTDOERFQNFKTLHMDFOKYKMWLEEN 840
QY 841 ERMNOEIVNLSKEAQKFDSSLGALKTSLSYKTOELQKTRQVQBRNEMEQLEKQLENRD 900
DB 841 ERMNOEIVNLSKEAQKFDSSLGALKTSLSYKTOELQKTRQVQBRNEMEQLEKQLENRD 900
QY 901 SPLQTVREKTLITEKLOQTLLEEVKTLTQEKDDLKQLESQIERDQKSDIHDTVMNNI 960
DB 901 SPLQTVREKTLITEKLOQTLLEEVKTLTQEKDDLKQLESQIERDQKSDIHDTVMNNI 960
QY 961 DTQQLRNALLESKQHOETINTLKSITSEEVSRNLHMEENTGETKDFQOMVGIDKKQD 1020
DB 961 DTQQLRNALLESKQHOETINTLKSITSEEVSRNLHMEENTGETKDFQOMVGIDKKQD 1020
QY 1021 LEAKNTQTLTADVKNDEIIEQORKIFSLIQEKNELQOMLESVIAEKEQKTDLKENIEMT 1080
DB 1021 LEAKNTQTLTADVKNDEIIEQORKIFSLIQEKNELQOMLESVIAEKEQKTDLKENIEMT 1080
QY 1081 IENQBELRLGDELKQKQOEIVAQEKNHAIKEGELSRTCDELARVEEKLKESQOLQEQK 1140
DB 1081 IENQBELRLGDELKQKQOEIVAQEKNHAIKEGELSRTCDELARVEEKLKESQOLQEQK 1140
QY 1141 QQLNVQSEMSEMOKKINEIENLNKELNKLTLTLEHMETERLELAQKLNENYEVKSYTK 1200
DB 1141 QQLNVQSEMSEMOKKINEIENLNKELNKLTLTLEHMETERLELAQKLNENYEVKSYTK 1200
QY 1201 ERKVLKELQKSFETERDHLRGVIRIEATGQTKELKIAHKLHEQHTIDELRRSVSE 1260
DB 1201 ERKVLKELQKSFETERDHLRGVIRIEATGQTKELKIAHKLHEQHTIDELRRSVSE 1260
QY 1261 KTAQIINTQDLKSHTKLQEEIPVLHBEQELLPNVKVSETQETWNELELLTEQSTTKDS 1320
DB 1261 KTAQIINTQDLKSHTKLQEEIPVLHBEQELLPNVKVSETQETWNELELLTEQSTTKDS 1320
QY 1321 TTLARIEMERLRLNKEKFOESQEEIKSLTKERDNLKTIKEALEVHKDQKHEIRETLAKIQ 1380
DB 1321 TTLARIEMERLRLNKEKFOESQEEIKSLTKERDNLKTIKEALEVHKDQKHEIRETLAKIQ 1380
QY 1381 ESQSQOESLNKKEKNDNTTKIVSEMOFKPKDSALLRIEIMLGLSKRLQSHDEMKS 1440
DB 1381 ESQSQOESLNKKEKNDNTTKIVSEMOFKPKDSALLRIEIMLGLSKRLQSHDEMKS 1440
QY 1441 AKEXDDLOQLQVLESQSDQKLENKEIVAKHLETEELKVAHCCCKQOETINELRVNL 1500
DB 1441 AKEXDDLOQLQVLESQSDQKLENKEIVAKHLETEELKVAHCCCKQOETINELRVNL 1500
QY 1501 SEKETEISTIQOLFAINDKLNQKIQEYKEEQOLNIQISVQBNVNLKOPKEHRKAK 1560
DB 1501 SEKETEISTIQOLFAINDKLNQKIQEYKEEQOLNIQISVQBNVNLKOPKEHRKAK 1560
QY 1561 DSALQSIESKMLNLTNRQLQESQEEIQIMIKEEMKRVQEQALQIERDQKLENKEIVAKM 1620
DB 1561 DSALQSIESKMLNLTNRQLQESQEEIQIMIKEEMKRVQEQALQIERDQKLENKEIVAKM 1620
QY 1621 KESQSEKEYOFLKMTAVNETQKQVCEIHLKEFQETQKMLNENIETENIILTQILLHENLEE 1680

Db 1621 KESQKEYQFLKMTAVNTQKMCIEIHLKEQFETQKLNENIETENIRLTQILHENLEE 1680
 QY 1681 MRSVTKERDDLRSEETLKVVERDQKLENRETITRDLEKQEBLKIVHMLKEHETIDKL 1740
 Db 1681 MRSVTKERDDLRSEETLKVVERDQKLENRETITRDLEKQEBLKIVHMLKEHETIDKL 1740
 QY 1741 RGVSEKTNESINMQKDLHSNDALKAQDLKIQEBELRIAHMLKEQOETIDKLRGIVSEK 1800
 Db 1741 RGVSEKTNESINMQKDLHSNDALKAQDLKIQEBELRIAHMLKEQOETIDKLRGIVSEK 1800
 QY 1801 TDKLSNMQKDLHSNAKLQKIELKANEHQITLTKKDVNNTQKQKSEMEQLKKIQKQK 1860
 Db 1801 TDKLSNMQKDLHSNAKLQKIELKANEHQITLTKKDVNNTQKQKSEMEQLKKIQKQK 1860
 QY 1861 LTLKLEIENLNAQELHENLEEMKSVKMDNRNRVEETLKLERDQKESLOETKARDL 1920
 Db 1861 LTLKLEIENLNAQELHENLEEMKSVKMDNRNRVEETLKLERDQKESLOETKARDL 1920
 QY 1921 EIQQELKTARMLSKHETVDKLRKISEKTIQISDIQKDLKSKDELQKKIQELQKKE 1980
 Db 1921 EIQQELKTARMLSKHETVDKLRKISEKTIQISDIQKDLKSKDELQKKIQELQKKE 1980
 QY 1981 QLLRVKEDVNMSHKINEMEQKKQEPNVLCKCENDNFQLTTKLHESLEEIRIVAKERD 2040
 Db 1981 QLLRVKEDVNMSHKINEMEQKKQEPNVLCKCENDNFQLTTKLHESLEEIRIVAKERD 2040
 QY 2041 ELRRIKESLKMEDQFIATLRMIARDNRHONQVKEPKLLSDGOQHLMSLEKCSRIKE 2100
 Db 2041 ELRRIKESLKMEDQFIATLRMIARDNRHONQVKEPKLLSDGOQHLMSLEKCSRIKE 2100
 QY 2101 LKRYSEMDDHYECLNRLSLDLKEIEFHRIMKGLKYVLSYVTKIEEQHECINKEMDF 2160
 Db 2101 LKRYSEMDDHYECLNRLSLDLKEIEFHRIMKGLKYVLSYVTKIEEQHECINKEMDF 2160
 QY 2161 IDEVEKQKELLIKIQLHQDCDVPSELRDLKLNQMDLHIEILKDFSEFPSPKTF 2220
 Db 2161 IDEVEKQKELLIKIQLHQDCDVPSELRDLKLNQMDLHIEILKDFSEFPSPKTF 2220
 QY 2221 QQVLSNRKEMTOFLEBWLNRFDIEKLNKGIQKENDRICOVNFFNRRIIATMNESTEF 2280
 Db 2221 QQVLSNRKEMTOFLEBWLNRFDIEKLNKGIQKENDRICOVNFFNRRIIATMNESTEF 2280
 QY 2281 ERSATISKWEQDLKSLKEKNEKLFKNYQTLKTSLSAGQVNPFTQDNKNPHVTSRATQ 2340
 Db 2281 ERSATISKWEQDLKSLKEKNEKLFKNYQTLKTSLSAGQVNPFTQDNKNPHVTSRATQ 2340
 QY 2341 TTEKIRELENSLHEAKESAMHESKIIKQKLEVTNDIIAKQKVHESNCKLEKTKET 2400
 Db 2341 TTEKIRELENSLHEAKESAMHESKIIKQKLEVTNDIIAKQKVHESNCKLEKTKET 2400
 QY 2401 IQVLQDKVALGAKPYKEETEDLKMVLKVIDLEKMKNAKEFEKISATKATVQKEVIRL 2460
 Db 2401 IQVLQDKVALGAKPYKEETEDLKMVLKVIDLEKMKNAKEFEKISATKATVQKEVIRL 2460
 QY 2461 LRENLRSSQQAQDTSVISEHTDPQSPNKLPTCGGSGIVQNTKALILKSEHRLKEISK 2520
 Db 2461 LRENLRSSQQAQDTSVISEHTDPQSPNKLPTCGGSGIVQNTKALILKSEHRLKEISK 2520
 QY 2521 LKQOQNEQLIKQNELLSNNQHLNNEVKTWKERTLKRHAHQVTCNSPKSPKVTGTASK 2580
 Db 2521 LKQOQNEQLIKQNELLSNNQHLNNEVKTWKERTLKRHAHQVTCNSPKSPKVTGTASK 2580
 QY 2581 KQITPSQCKERNLQDPVKESPKSCFFDPSRSKSLPSPHPVRYFDNSSGLCPEVQAGAE 2640
 Db 2581 KQITPSQCKERNLQDPVKESPKSCFFDPSRSKSLPSPHPVRYFDNSSGLCPEVQAGAE 2640
 QY 2641 SVDSQFGPWHASSGKDVPECKTQ 2663
 Db 2641 SVDSQFGPWHASSGKDVPECKTQ 2663

RESULT 2

GOB1_HUMAN

ID AC Q14739; Q14398; STANDARD; PRT; 3259 AA.
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Golgi autoantigen, golgin subfamily B member 1 (Giantin) (Macroglolin)
 DE (Golgi complex-associated protein, 372-kDa) (GCP372).
 GN GOLGB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94187728; PubMed=7511208;
 RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
 RA Renz M.;
 RT "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane
 RT protein (giantin).";
 RL Mol. Cell. Biol. 14:2564-2576(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94257116; PubMed=8198703;
 RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
 RA Renz M.;
 RT "Macroglin -- a new 376 kD Golgi complex outer membrane protein as
 RT target of antibodies in patients with rheumatic diseases and HIV
 RT infections.";
 RL J. Autoimmun. 7:67-91(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95100974; PubMed=7802676;
 RA Sonda M., Misumi Y., Fujiwara T., Nishioka M., Ikehara Y.;
 RT "Molecular cloning and sequence analysis of a human 372-kDa protein
 RT localized in the Golgi complex.";
 RL Biochem. Biophys. Res. Commun. 205:1399-1408(1994).
 CC -!- FUNCTION: May participate in forming intercisternal cross-bridges
 CC of the Golgi complex.
 CC -!- SUBUNIT: Homodimer; disulfide-linked.
 CC -!- SUBCELLULAR LOCATION: Golgi; membrane-associated.
 CC -!- DISEASE: Antigen in chronic rheumatoid arthritis and in the
 CC autoimmune disease Sjogren's syndrome.
 CC -!- SIMILARITY: Belongs to the golgin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; X75304; CAA53052.1; --
 CC EMBL; D25542; BAA05025.1; --
 CC PIR; A56539; A56539.
 CC PIR; I52300; I52300.
 CC Genew; HGNC:4429; GOLGB1.
 CC MIM; 602500; --
 CC DR GO; GO:000139; C:Golgi membrane; TAS.
 CC DR GO; GO:0005795; C:Golgi stack; TAS.
 CC DR GO; GO:0016021; C:integral to membrane; TAS.
 CC DR GO; GO:0007030; P:Golgi organization and biogenesis; TAS.
 CC DR Golgi stack; Antigen; Coiled coil; Transmembrane.
 CC KW DOMAIN 1 3235
 CC FT TRANSMEM 3236 3256
 CC FT DOMAIN 3257 3259
 CC FT DOMAIN 48 593
 CC FT DOMAIN 677 1028
 CC FT DOMAIN 1062 1245
 CC FT DOMAIN 1301 1779
 CC FT DOMAIN 1828 3185
 CC FT DOMAIN 2420 2423
 CC FT DOMAIN 2993 2996
 CC FT POLY-SER.


```

RESULT 3
GOA4_HUMAN
ID   GOA4_HUMAN          STANDARD;          PRT:  2230 AA.
AC   Q13439, Q13270; Q13654; Q14436;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DE   Golgi autotransmembrane protein, golgin subfamily A member 4 (Trans-Golgi p230) (256
DE   kDa golgin) (Golgin-245) (72.1 protein).
DN   GOLGA4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=96215236; PubMed=8628529;
RA   Erlich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.-H.;
RT   "Molecular characterization of trans-Golgi p230: a human peripheral
RT   membrane protein encoded by a gene on chromosome 6p12-22 contains
RT   extensive coiled-coil alpha-helical domains and a granin motif.";
RL   J. Biol. Chem. 271:8328-8337(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Seelig H.P.;
RL   Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN   [3]
RP   SEQUENCE OF 131-2230 FROM N.A.
RX   TISSUE=Placenta;
RA   MEDLINE=96125113; PubMed=8537393;
RT   Fritzler M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan E.K.L.;
RT   "Molecular characterization of golgin-245, a novel Golgi complex
RT   protein containing a granin signature.";
RL   J. Biol. Chem. 270:31262-31268(1995).
RN   [4]
RP   SEQUENCE OF 524-672 FROM N.A.
RX   TISSUE=Gastric fundus;
RA   Balague C.;
RL   Thesis (1994), Instituto municipal de investigacion medica, Spain.
CC   -!- FUNCTION: May play a role in vesicular transport from the trans-
CC   Golgi.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic, peripheral membrane protein
CC   associated with the trans-Golgi network.
CC   -!- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=4;
CC   Comment=Additional isoforms seem to exist;
CC   Name=1;
CC   IsoId=Q13439-1; Sequence=Displayed;
CC   Name=2;
CC   IsoId=Q13439-2; Sequence=VSP_004272, VSP_004273;
CC   Name=3;
CC   IsoId=Q13439-3; Sequence=VSP_004274;
CC   Name=4;
CC   IsoId=Q13439-4; Sequence=VSP_004275;
CC   -!- DOMAIN: Extended rod-like protein with coiled-coil domains.
CC   -!- DISEASE: Antibodies against GOLGA4 are present in sera from
CC   patients with Sjogren's syndrome (SS) [MIM:270150]. Sera from
CC   patients with SS often contain antibodies that react with normal
CC   components of the Golgi complex.
CC   -!- DISEASE: Antibodies against GOLGA4 are found in sera from
CC   hepatitis B patients.
CC   -!- SIMILARITY: Belongs to the golgin family.
CC   -!- SIMILARITY: Contains 1 GRIP domain.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/
CC   or send an email to license@isb-sib.ch).
CC   -----
CC   EMBL; U41740; AAC50434.1; -

```



```

Db 2666 EKKQREKKEKRPQDVEVLKTTTTELFHNSNEESGFFNELEALRAB--SVATKAELASY-- 2720
QY 2022 TKKLSHLSLEIRIVAKE-----RDLRLIKESLK_MERDQFIATLREMIARD 2067
Db 2721 -KEAEKLOE-ELLVKETNTMTSLQDLSQVRDHLAEAKELSLILEKEDTEVOESKACM 2778
QY 2068 RQNHQVPEKRLSLSDGQHLMESLREKCSRIKELLKRYSEMDDHYECLNRLSLDLLEKHE 2127
Db 2779 FELPIKLSIASQDGTGLKISSNQTPQI--LVKNAG-----IQINLQSECS 2825
QY 2128 FHRIMKKLVYLVYTKIKKEQHECINKFEMDFIDEVEK--OKELLIKIQIHLQDCDVPSP 2186
Db 2826 SEEVTEIISQTEKIEKQELHAAEILDMESRHSITETITLKRHYAVQLLECEGTLKA 2885
QY 2187 ELRDLK-----NONDLHTEETLKDFSES---EFP 2214
Db 2886 VIQCLRSKRVGFYFNMCFSTLCSGSDWGQGIYILTHSQGFDI-ASEGREGESSATDSFP 2944
QY 2215 -----SIKTEFOQVLS-----NRKEMTQFLEWLNTR-----FDIEKLKNGIQK 2253
Db 2945 KKIKGLLRVHNEGQVLSLTSPSYDGDHDSIQOVSEPWLEERKAYINTISSIKOLITK 3004
QY 2254 -ENDRICQVNN-----FFNRIITAINNESTE----- 2278
Db 3005 MQLQREAVDYDSSQSHESFSDWRGELLALQVFEERSVILLAAFRTELTAFTGTTDAVL 3064
QY 2279 ---PEERSATISKWEODLSKKEKNEK-LFNQVOTLTKTSLASGAQVNPPTQDNKNP--- 2331
Db 3065 LNCLEQRIQGGVEYQAMECLQADRRSLSEIQLHAQWNGKIIYLKREQSEKESQSE 3124
QY 2332 -----HVTSRAT-----QUTTEK---IRELENSLHEAK-----ES 2358
Db 3125 LLEYNIQKQSQMLEQVVELSSMKDRATELQELQSEKVMVAELKSELAQTKLETTLK 3184
QY 2359 AMHKEKIIKMKQ-----ELEVTNDIATKLOAKVHESKNCLKEKKEITIQVLQDKVAL 2410
Db 3185 AQHKLKLEALFRLEVKDTEHLNDTLASEQKSGRELQWALEKEKA-----KLGR 3237
QY 2411 GAKPYKEEIDLMKVLKIDLEKMKNAKEPEKESATKATVEYQKVEVIRLLRENLSQQ 2470
Db 3238 SEEDKELELDKFSL---ESQKORNQ-----INLLLEQKQLINESQKIESQRM 3286
QY 2471 AQDTSVISEHTDPOPSKNPLFCGGSGIVQNTKALIKSEHIREKEISKLKQONE----- 2526
Db 3287 LYDAQLSSE---QGRNLEL-----QVLLSEKVRIREMSSTLDRERLHAQ 3329
QY 2527 -----OLIKQKNELLNNQHLSENVKTKERTLK-----RE 2557
Db 3330 LQSSDGTGQSRPPLPSEDLKLEKQLEKHSRIVELLNETEKYKIDSLQTRQMEKDRQ 3389
QY 2558 AHKQVTCENSPKSPKVTGTASKKQITPSQCKERNLQDPVPKESPKCFPDSRSKSLPSP 2617
Db 3390 VHRKYL-----QTEQANTEQKK-MHELSQKVEDLQRLQLEKROQVYKLDLEGQRLQ-- 3441
QY 2618 HPVRYFDSNLSGLCEVQNAESVD-----SQCPWIASGK 2655
Db 3442 -----GIMQEFQKQELERBEKESRRILYQNLNEPTTWSLTSR 3480

RESULT 5
CENF_HUMAN
ID AC P49454; Q13171; Q13246; PRT; 3210 AA.
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE CENP-F kinetochore protein (Centromere protein F) (Mitosis) (AH antigen).
GN CENP-F.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```

```

RN RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=95348175; PubMed=7542657;
RA Liao H., Winkfein R.J., Mack G., Rattner J.B., Yen T.J.;
RT "CENP-F is a protein of the nuclear matrix that assembles onto
RL kinetochores at late G2 and is rapidly degraded after mitosis.";
RN RP J. Cell Biol. 130:507-518(1995).
RN RP [2]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=95379848; PubMed=7651420;
RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B.,
RN Jones D., Yang-Peng T.L., Lee W.-H.;
RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein
RL that is specifically involved in mitotic-phase progression.";
RN RP Mol. Cell. Biol. 15:5017-5029(1995).
RN RP [3]
RN RP SEQUENCE OF 2194-3210 FROM N.A.
RX MEDLINE=95336446; PubMed=7612011;
RA Li Q., Ke Y., Kapp J.A., Fertig N., Medsger T.A. Jr., Joshi H.C.;
RT "A novel cell-cycle-dependent 350-kDa nuclear protein: C-terminal
RL domain sufficient for nuclear localization.";
RN RP Biochem. Biophys. Res. Commun. 212:220-228(1995).
RN RP [4]
RN RP CHARACTERIZATION
RX MEDLINE=95370296; PubMed=7642639;
RA Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley W.R., Lee W.-H.;
RT "The C terminus of mitosis is essential for its nuclear localization,
RL centromere/kinetochore targeting, and dimerization.";
RN RP J. Biol. Chem. 270:19545-19550(1995).
RN RP [5]
RN RP CHARACTERIZATION
RX MEDLINE=9437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RL interactions with the kinetochore proteins CENP-F and hBUBR1.";
RN RP J. Cell Biol. 143:49-63(1998).
RN RP [6]
RN RP FARNESYLATION
RX MEDLINE=20459117; PubMed=10852915;
RA Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L.,
RN Bishop W.R., Kirschmeier P.;
RT "Farnesyl transferase inhibitors block the farnesylation of CENP-E
RL and CENP-F and alter the association of CENP-E with the
RN microtubules.";
RN RP J. Biol. Chem. 275:30451-30457(2000).
RN RP [7]
RN RP FUNCTION: Probably required for kinetochore function, involved in
CC chromosome segregation during mitosis. Interacts with
CC retinoblastoma protein (RB), CENP-E and BUBR1.
CC [8]
CC SUBUNIT: Homo- or heterodimer.
CC [9]
CC SUBCELLULAR LOCATION: Nuclear matrix (but not in the nucleolus),
CC reorganization to the kinetochore/centromere (coronal surface of
CC the outer plate) and the spindle during mitosis.
CC [10]
CC DEVELOPMENTAL STAGE: Gradually accumulates during the cell cycle.
CC [11]
CC PTM: Hyperphosphorylated during mitosis.
CC [12]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC [13]
CC EMBL; U19769; AAA82889.1; -
CC EMBL; U30872; AAA82935.1; -
CC EMBL; U25725; AAA86889.1; -
CC PIR; PC4035; PC4035
CC Genew; HGNC:1857; CENPF.
CC GK; P49454; -
CC MIM; 600236; -
CC GO; GO:0005699; C:kinetochore; TAS.
CC GO; GO:0005634; C:nucleus; TAS.

```


1786 VHQICDQAQDNLNLDIEKITETGAVKPTGCSGSGSPDTNYPGDEKTKQSSCSISBL 1845
 1302 -----QETWNELELLTEOSTTKOSTTL-----A 1324
 1846 SFGSPNALVPMDFLNGQDIEHNLQFVKETNSNENRLLHVEDRDRKVESLNNEMKELDS 1905
 1325 RIEMERLRNNEKQF---ESQEEIKSLTKERDNLTKIKEALEVHKDQ----- 1368
 1906 KUHLQEVQWTKTEACIELEKIVGELKKENDSLSEKLEVFSCDHQELLQFVETSEGLNSD 1965
 1369 -----KEHIRETLAKIOESQSQOSQOSLNKMKENDNETTKIVSEMEQPKPKDQALLRI 1419
 1966 LEMHADKSPREDIGDVAKNVDSW---KERFLDV---ENELSRIRSEKASIE-HEALYLEA 2019
 1420 EIMGLGLKPLQSHDEMKS-----VAKKDDIOLRQEVLOSSES---DOLKEN 1464
 2020 DLEVVQTEKLCLEKDNENKQKVIIVCLLEELSVVTSERNQURGELDTMSKKTALDQUSEK 2079
 1465 IKEIVAKHLTEEE---LKVHACCLKEQETINELRVLNLSKETEISTIOKQLEAIND 1519
 2080 MKE-KTQELSHOSECHLCIOVAFAVEKTELLQTLSSDVSELLKDKTHLQELQSLBK 2138
 1520 -----KLQNKIOEIEKEBQNLNKOISEVQENNELKQFKEHRKADS-ALQSI 1568
 2139 DSQALSITKCELENOIAQL-NKEKELLVKESESLOARLSE---SDYEKLNVSKALEAAL 2193
 1569 SKMLETNRLQESQEEIOMIK-----EKEEMKRVQALOI-----ERDQKENTK 1614
 2194 VEKGEFALRLSSTOEEVHQHARGIEKLRVIRTEADKKQLHIAEKLKERENDSLKQV 2253
 1615 EIVAKMESQSEKQFQKMTAVNETQKMCIEHLKEQFETQ-----KLNLENIFEN 1667
 2254 NLERELQSENOQ-----ELVILDAENSKAHEVETLKTQIEEMARSLKVFELDLVLRSEK 2308
 1668 IRLTQILHEN-----LEEMRSVTKERD---DLRSVET---LKVRODLKENLRET 1712
 2309 ENLTQIQEKQQLSELDKLSSPKSLLEEQAQBIQIKESKTAVEMLQNLKE-LNEA 2367
 1713 ITRDLKQEBELKIVMHL-----KEHQ---ETIDKLRG-IVSEKTNIEISNMQ---KDLHSN 1762
 2368 VAALCGDQEIWKATEQSDLPPIEEHQLRNSIEKURARLEADEKKQLCVLQKKESEHHA 2427
 1763 DALKAADLKIOBELRIAHMLHKEQOETIDKLRGIVSEKTDKLSNMOKOLENSNAKLQEBKI 1822
 2428 DLLKRVENLRELEIA---RTNQE-----HAALRAENSKEGVEVETLKAKIEGMT 2473
 1823 QELKANERHLITLKKD---VNETQKK-----VSEBOLKQIKQDOSL 1861
 2474 QSLRGLDVTVIRSEKEDLTNELQKEBRISELEIINSFPENILOKEKEQKVQMKES 2533
 1862 TLKSL---EIEMLN-LAQELHENLEEMKSVMKERNLRRVETTLERDLKESIQETKA 1917
 2534 TAMEMLOTQLKELNERNVAALHNDQACGA---KEQLSSQVECLEKAKQLLQGLDEAKN 2590
 1918 RDLTEQELKLTARMLSKHEKHTVDKLRKXISEKTIQISDIQDKLDSKD----- 1966
 2591 NVIVLQSSVNGLIQFVEDGQKLEKDEBISRLKNGIQOQEQVLVSKLSQVEGHEHQLWKEQ 2650
 1967 -----ELQKKIQELQKELQLLRVVEDVNMHKKI-NEMEQLKQFEPNYLCKEM 2016
 2651 NLELRNLATLEQKIQVLQSKNASLODTTLEVLQSSYKNLENELELT-----M 2698
 2017 DNFQTLKLHESLEBRIIVAKERDELRLRIKESLKMERDQFIATLRMIARDRONHVKPE 2076
 2699 DKMSFVEKNV-----KMTAKETELOREHEMA-----QKTALQOE-----ELSGE 2738
 2077 KRLSDGQOHLMBLSREKCSRIKELKRYSEMDDHYECNLNRLSLDLKEKEIEFRIMKLLK 2136
 2739 KNPRLAGLQLLLEIKSSQDKLQELTLENSELKSLDCMHK--DQVEKE----- 2785
 2137 YVLSYVTKIKEQHECINKFEMPFDIVBQKQELLIKIHLQOQCDVPS-RELRLDLKLNQ 2195
 2786 -----GKVRSE-----IABYQLR-LHEAKKHQALLDNTKQVEVEIQTYRE-----KLTS 2830

2196 NMDLHIEELKDFSESEFPIKTEFOVLNRSKEMTQFLEEWLNTFRDIEKLKNGIQKEN 2255
 2831 K-----ECLSS-QKDEIDLKSSKEELNNSLKATTOILEELKTKTMDNLKYVNLKKN 2884
 2256 DRICOVNFFNRIITAINNESTEFERSATI-SKEWQDLSKLKKEKNEKLFKQYQTLKTS 2315
 2885 ERA-----QGMKLLIKSCQLEKEEILQKELSQ-LQAAQEK----- 2921
 2316 ASGAQVNPQTQDNKNPHVTSRATQLTTEKIRELENSLHEAKESA---MHKESKLIIMQKE 2372
 2922 ----QKTGTWMDTK-----VDELATTE-IKELKTELEKTKDEADYIDKYCSLLISHEK 2969
 2373 LEVNDIIIAKLQAKVHESNKLCKETKETTQ---VLQDKVALGAKPYKEEIED-LKMKLVKI 2429
 2970 LEK-----AKEMLETOVAHLCSQOSKQDSGRGSPLLGPVVPSPPIPSVTEKRLSSGQNK 3024
 2430 DLEKMKNAKEFEKIEISATKATVE-YQKEVIRLRENLRRSQOQDTSVISEHTDQPSNK 2488
 3025 SGRQRSSGIWENGPGTTPATPESFSKSKKAVMSGI---HPAEDT---EGTEFEPEGL 3077
 2489 PLTCGGSG-----IVQNTKALILKSEHRLK-EISKLQOQNEOLIKQKNELLS 2537
 3078 PEVVKKGFDIPTKTSPIYLRRTTMARTSPRLAAOKLALSPLSLKENLAESKKTAG 3137
 2538 NNQHLNNEVTKWERTLKREAHKQVTCENSGPKPKVTGTASKKKQITPSQCKERNLODPV 2597
 3138 GSR--SOKVY-----AQRSPVDSSTILREP--TTKSPVNVNLPERS 3175
 2598 PKSPKSCFFDSRSKSLPSPHPVRYFNDSSLGLCPFVQNGAESVDQPGFWHASSGKV 2657
 3176 PTDSPREGLRVKGRGLVSP-----KAGLESKGS----- 3205
 2658 PECKTQ 2663
 3206 -NCKVQ 3210

RESULT 6
 CEP2 HUMAN
 ID CEP2 HUMAN STANDARD; PRT: 2442 AA.
 AC Q9VW73; O14812; O60588; Q9H450;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Centrosomal protein 2 (Centrosomal Nek2-associated protein 1) (C-NAP1)
 DE (Centrosome protein 250) (Centrosome associated protein CEP250).
 GN CEP2 OR CNAP1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RC SEQUENCE FROM N.A. (ISOFORM 1), AND AUTOIMMUNE DISEASE.
 RP TISSUE=Cervical carcinoma;
 RX MEDLINE=98165428; PubMed=9506584;
 RA Mack G.J., Rees J., Sandblom O., Balczon R., Fritzler M.J.,
 RA Rattner J.B.;
 RT "Autoantibodies to a group of centrosomal proteins in human autoimmune
 RT sera reactive with the centrosome.";
 RL Arthritis Rheum. 41:551-558 (1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), INTERACTION WITH NEK2,
 RP AND SUBCELLULAR LOCATION DURING THE CELL CYCLE.
 RC TISSUE=Placenta;
 RX MEDLINE=98311641; PubMed=9647649;
 RA Fry A.M., Mayor T., Meraldi P., Stierhof Y.-D., Tanaka K., Nigg E.A.;
 RT "C-Nap1, a novel centrosomal coiled-coil protein and candidate
 RT substrate of the cell cycle-regulated protein kinase Nek2.";
 RL J. Cell Biol. 141:1563-1574 (1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=21638749; PubMed=11780052;

RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Clegg S.,
RA Clegg S., Copley V.E., Collier R.E., Clark L.N., Clark S.Y., Clee C.M.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Hunt A.R., Hunt P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Johnson C.M., Johnson D.,
RA Leharshaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.A., Moore M.J.F., Mullikin J.C., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Thorpe A.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.,
RA "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Placenta;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Villalón D.K., Muzny D.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RP PHOSPHORYLATION, AND INTERACTION WITH NEK2 AND PPPICA.
RX MEDLINE=21060765; PubMed=10880350;
RA Helps N.R., Luo X., Barker H.M., Cohen P.T.W.;
RT "NIMA-related kinase 2 (Nek2), a cell-cycle-regulated protein kinase
RL localized to centrosomes, is complexed to protein phosphatase 1.";
RL Biochem. J. 349:509-518(2000).
[6]
RP PHOSPHORYLATION DURING CELL CYCLE.
RX MEDLINE=22135747; PubMed=12140259;
RA Mayor T., Hacker U., Stierhof Y.-D., Nigg E.A.;
RT "The mechanism regulating the dissociation of the centrosomal protein
RT C-Nap1 from mitotic spindle poles.";
RL J. Cell Sci. 115:3275-3284(2002).
CC -!- FUNCTION: Probably plays an important role in centrosome cohesion
CC during interphase.
CC -!- SUBUNIT: Monomer and homodimer (Probable). Forms a complex in
CC vitro with both Nek2 kinase and the PPP1C catalytic subunit of
CC protein phosphatase 1 (PP1).
CC -!- SUBCELLULAR LOCATION: Component of the core centrosome. In
CC interphase cells, it specifically associates with the proximal
CC ends of both mother and daughter centrioles. Associates with the

CC centrosome in interphase cells. In mitotic cells, it dissociates
CC from the mitotic spindle poles. At the end of cell division, it
CC reaccumulates at centrosomes.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=Q9BV73-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9BV73-2; Sequence=VSP_007372;
CC Name=3;
CC IsoId=Q9BV73-3; Sequence=VSP_007370, VSP_007371;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Ubiquitously and weakly expressed.
CC -!- PTM: Differentially phosphorylated during cell cycle.
CC Phosphorylation may regulate association/dissociation from
CC centrosome. During M phase of mitosis, C-terminal part is
CC phosphorylated by NEK2, suggesting that it may trigger the
CC dissociation from the mitotic centrosome. It is dephosphorylated
CC in vitro by the PP1 phosphatase.
CC -!- DISEASE: Antibodies against CEP2 are present in sera from patients
CC with autoimmune diseases that developed autoantibodies against
CC centrosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; AF022655; AAC06349.1; -;
CC EMBL; AF049105; AAC07988.1; -;
CC EMBL; AL121586; CAB89415.1; -;
CC EMBL; BC001433; AAH01433.1; -;
CC Genew; HGNC:1859; CEP2.
CC Cell cycle; Coiled coil; Phosphorylation; Alternative splicing;
CC Polymorphism.
CC DOMAIN 95 158
CC DOMAIN 244 352
CC FT COILED COIL (POTENTIAL).
CC FT COILED COIL (POTENTIAL).
CC FT COILED COIL (POTENTIAL).
CC FT COILED COIL (POTENTIAL).
CC FT COILED COIL (POTENTIAL).
CC FT POLY-LGU.
CC FT GLN/GLU-RICH.
CC FT V -> F (in isoform 3).
CC FT /FTid=VSP_007370.
CC FT Missing (in isoform 3).
CC FT /FTid=VSP_007371.
CC FT Missing (in isoform 2).
CC FT /FTid=VSP_007372.
CC FT Q -> H (in dbSNP:2296403).
CC FT /FTid=VAR_015649.
CC FT L -> I (IN REF. 1).
CC FT E -> A (IN REF. 1).
CC FT H -> L (IN REF. 1).
CC FT D -> E (IN REF. 1).
CC FT E -> D (IN REF. 2).
CC FT S -> I (IN REF. 1).
CC FT E -> A (IN REF. 1).
CC FT EVTK -> DEPQ (IN REF. 1).
CC FT Q -> H (IN REF. 1).
CC FT H -> L (IN REF. 1).
CC FT L -> P (IN REF. 1).
CC FT CONFLICT 120 120
CC FT CONFLICT 136 136
CC FT CONFLICT 365 365
CC FT CONFLICT 372 372
CC FT CONFLICT 509 509
CC FT CONFLICT 552 552
CC FT CONFLICT 757 757
CC FT CONFLICT 784 787
CC FT CONFLICT 1153 1153
CC FT CONFLICT 1246 1246
CC FT CONFLICT 1513 1513
CC FT CONFLICT 2082 2082
CC FT CONFLICT 2345 2345
CC FT CONFLICT 2442 2442
CC FT SEQUENCE 2442 AA; 281135 MW; BC2B8A36E07B8272 CRC64;
CC

Query Match 7.7%; Score 1032; DB 1; Length 2442;
Best Local Similarity 20.7%; Pred. No. 1.1e-20;
Matches 562; Conservative 539; Mismatches 944; Indels 674; Gaps 112;

Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
"The genome sequence of *Drosophila melanogaster*.";
Science 287:2185-2195 (2000).
[2]
SEQUENCE OF 980-2779 FROM N.A.
RP STRAIN=Berkeley; TISSUE=Embryo;
RC MEDLINE=20530668; PubMed=12537569;
RX MEDLINE=22426066; PubMed=12537569;
RA George R.M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George L.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A *Drosophila* full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).
[3]
FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RP MEDLINE=20530668; PubMed=11076973;
RX MEDLINE=20530668; PubMed=11076973;
RA Sisson J.C., Field C., Ventura R., Royou A., Sullivan W.;
RT "Lava lamp, a novel peripheral golgi protein, is required for
RT *Drosophila* melanogaster cellularization.";
RL J. Cell Biol. 151:905-918 (2000).
CC -!- FUNCTION: Lva and spectrin may form a Golgi-based scaffold that
CC mediates interaction of Golgi bodies with microtubules and
CC facilitates Golgi-derived membrane secretion required for the
CC formation of furrows during cellularization.
CC -!- SUBUNIT: Interacts with CLIP-190 and spectrin separately.
CC -!- SUBCELLULAR LOCATION: lva-alpha-spectrin and lva-CLIP-190
CC complexes are found at the Golgi.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AE003429; AAF45910.1; -;
DR EMBL; AY118638; AAM50007.1; ALT_INIT.
DR FlyBase; FBgn029688; lva.
KW Developmental protein; Golgi stack; Coiled coil.
FT DOMAIN 52 85 COILED COIL (POTENTIAL).
FT DOMAIN 141 175 COILED COIL (POTENTIAL).
FT DOMAIN 220 607 COILED COIL (POTENTIAL).
FT DOMAIN 659 716 COILED COIL (POTENTIAL).
FT DOMAIN 751 1733 COILED COIL (POTENTIAL).
FT DOMAIN 1785 1863 COILED COIL (POTENTIAL).
FT DOMAIN 1941 2433 COILED COIL (POTENTIAL).
FT DOMAIN 2504 2544 COILED COIL (POTENTIAL).
FT DOMAIN 2600 2641 COILED COIL (POTENTIAL).
FT DOMAIN 1211 1211 E -> K (IN REF. 2).
FT CONFLICT 1424 1424 E -> S (IN REF. 2).
FT CONFLICT 1506 1506 E -> D (IN REF. 2).
FT CONFLICT 2034 2034 A -> V (IN REF. 2).
FT CONFLICT 2069 2069 G -> H (IN REF. 2).
FT CONFLICT 2133 2133 G -> E (IN REF. 2).
FT CONFLICT 2154 2154 Q -> E (IN REF. 2).

```

CC mediates interaction of Golgi bodies with microtubules and
CC facilitates Golgi-derived membrane secretion required for the
CC formation of furrows during cellularization.
CC
CC -! SUBUNIT: Interacts with CLIP-190 and spectrin separately.
CC -! SUBCELLULAR LOCATION: Lva-alpha-spectrin and Lva-CLIP-190
CC complexes are found at the Golgi.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@lsb-sib.ch).
CC
CC -----
CC DR EMBL; AE003429; AAP45910.1; -
CC DR EMBL; AY118638; AAM50007.1; ALT_INIT.
CC DR FlyBase; FBgn029688; lva.
CC KW Developmental protein; Golgi
CC
CC FT DOMAIN 52 85
CC FT DOMAIN 141 175
CC FT DOMAIN 220 607
CC FT DOMAIN 659 716
CC FT DOMAIN 751 1733
CC FT DOMAIN 1785 1863
CC FT DOMAIN 1941 2433
CC FT DOMAIN 2504 2544
CC FT DOMAIN 2600 2641
CC FT CONFLICT 1211 1211
CC FT CONFLICT 1211 1221
CC FT CONFLICT 1424 1424
CC FT CONFLICT 1506 1506
CC FT CONFLICT 2034 2034
CC FT CONFLICT 2069 2069
CC FT CONFLICT 2133 2133
CC FT CONFLICT 2154 2154
CC
CC
CC stack; Coiled coil.
CC COILED COIL (POTENTIAL).
CC COILED COIL (POTENTIAL).
CC COILED COIL (POTENTIAL).
CC COILED COIL (POTENTIAL).
CC COILED COIL (POTENTIAL).
CC COILED COIL (POTENTIAL).
CC COILED COIL (POTENTIAL).
CC COILED COIL (POTENTIAL).
CC COILED COIL (POTENTIAL).
CC E -> K (IN REF. 2).
CC A -> S (IN REF. 2).
CC E -> D (IN REF. 2).
CC A -> V (IN REF. 2).
CC Q -> H (IN REF. 2).
CC G -> E (IN REF. 2).
CC O -> E (IN REF. 2).

```


FT CONFLICT 2160 2160 E -> V (IN REF. 2).
 FT CONFLICT 2200 2200 L -> P (IN REF. 2).
 FT CONFLICT 2217 2217 A -> T (IN REF. 2).
 FT CONFLICT 2271 2271 E -> D (IN REF. 2).
 SQ SEQUENCE 2779 AA; 315897 MW; 1CB3965102018AEE CRC64;
 Query Match 7.2%; Score 959; DB 1; Length 2779;
 Best Local Similarity 20.2%; Pred. No. 1e-18;
 Matches 569; Conservative 528; Mismatches 951; Indels 768; Gaps 117;
 QY 152 LIREDVNRNVYVADLIEEYVYVYSEMALKWITKXSRHYGETYKMQRSRS-----HT 205
 Db 305 LVLRERLAELENVYDLETT-RCELELTTRARORNLRELEQOEKESRQSEAAHT 362
 QY 206 IFRMILE-SREKPEPNCESKVKVSHNLVYL-AGSERRAQTGAAGVRLKEGCVNRSFL 263
 Db 363 DAQVSABLAKOLOELTN-----QLADLOATNEELRQQAQAQKLVOTDE-----406
 QY 264 ILGOVLIKLSGQGVGFYNYRDSKLTIRLQNSLGGNPKTRI-ICTIIPVSFDETLTALQF 322
 Db 407 IVSORLEELRATTAQELLELOEQKSAMAAQNEELAEKTTTELNVNLNRLLEEKI-----461
 QY 323 ASTAKYMNTP-YNEVSTDEALRYRKEIMDKKOLEE-----VSETRQAAMEK-----373
 Db 462 ---AQSRKPLFLEDHSDSAAKQMDIQQIKLIDETNKANIKLCKCKQAEKKLQ 518
 QY 374 -----DOLAQLLEKLLQK-----VONEK-----IENL-----397
 Db 519 KFSQDQGOQQLASLALNEELOQRIAVLEDEKQWQLANMOEDDRQPEQSTESNPLQLE 578
 QY 398 TRMLVTSSTLLOQLKAKR-----KRVTWCLGKINKMKNVYADQFNI-----442
 Db 579 TIRLEBQLELOQALEALLSSSSSAESIEIVERHLECLQRRPASEGDAQEQKVHPP 638
 QY 443 -PTNIT--TKTHKLSINLLRIDESVCSDFVNTLDTLSEIWNPAKLLNQNIESE 499
 Db 639 GPSHVSELTQEQ-----TEEDSSGETLSQLE-----RLEFTQERGE-----678
 QY 500 LNSLRADYDNLVD-YEQLRTEKEBEMELKKEKDLDEFEALERTKKQDMQLIHEISN 558
 Db 679 -----VLDKLEQSAENLQOLARLESS--SSLQLLQREKQ-----LISSST 721
 QY 559 LKNLVKHEVYNODLE-----NELSKVVELLREKEDQTKKLOE 596
 Db 722 SSNLQESLSSMRSESVATLDAGEGPVLFKCEKSLKSLNSELEYKANDROAKFNV 781
 QY 597 YIDSQK-----ENIKMDLSYLESIEDPKMQKQTLFDAET-VALDAKRESAF-----LRS 646
 Db 782 ---SKLAKEAKNCHTQLSLELLHKVKEASTAVETVTVVAVTAPNGKALAEYEQUNA 838
 QY 647 ENLELK-----EKMKELATTY-----KOMENDIQLYQSOLEAKKMQVDL 686
 Db 839 QNAELKAVISRLOQELDELRESYPTEAPLAVGSDSQREDEILQLOQLEDAESLQAEQ 898
 QY 687 EKELQSAFNEITKLSLIDGKVPKOLLNLEKGITDLO-----KELN--KE 732
 Db 899 ROQIEQVQVQIKELRQ-----TEAQOLQVARQSAEITQLOQSEQPDOLLNSKENSHEKQ 954
 QY 733 VEENALREBEVILLBLKLPSEVERLKEIKQKSELHIITSEKDLSEVVHKBRSVQ 792
 Db 955 LEOQTRIRRE--LEARAESLEGELSITQIVAEQKOOLTESSESHALNKLMLQESQA 1012
 QY 793 GLLBELGKTD-----DLATTQSNVKSITDQSFQNFKTL-HMDFEQ 831
 Db 1013 EELRELRKEDDPOLRELRVSKSLVAQVRELTSSQETVDALNQIQEYQGLEHAHKEE 1072
 QY 832 KYKMWLEENRMNQEIWNLSKEAQKFDSSIGALKTELSTYKQTOLEQKRETVQBRLEMEQ 891
 Db 1073 QFNK-RELREKLKYALNKKRTQ-----DNADLEQKVQELTSQLOEQOE 1116
 QY 892 LKEQLENRDSPLQTVREKTLITEKLUQOTLEFVKLTQEQ-----KDDLKQLOES 940
 Db 1117 LVQKKEE-----VEREPIVDNHRVQLOQVQSKLNEDLKAKIHLNLRDALRLQKQ 1169

QY 941 LOIERDOL-----KSDIHDTVNMNIDTQEOLRNALESILKQOETINTLKSKEI SEEVSRLNH 996
 Db 1170 IQ-EQQLIQRDAELODA---NLVSKL-LRRRQEADEVFQLOQENSRLREISKLOE 1224
 QY 997 MEENTGETKDEFOQKMGVGDKKQDLKAKNTQTLTADVKNELIEBQOKKIFSLIOEKNELO 1056
 Db 1225 EHNLOQRVNE--EPTAVEDLRQLEAKSKF---EKSKELIKLRNATIQSIQ--RELQ 1276
 QY 1057 QMLESVIARKEQKLTDLKENIEMTIENQBELRLLGDELKKQOIVAEQKNNHAIKKEGE-- 1114
 Db 1277 QLOQDQDSEVHRNARAHEQLRLEKDAEITAL-----RQBIKLKLSRA-AGEGDDT 1329
 QY 1115 LSTCDRLAEVEKLEKSKOQLOEQKQOLNVQEMSEMOKKINEIENKLNKLNKELTL 1174
 Db 1330 ITKTSQLESQOQAEBSQVARELQQLRVOLTAQEQHAL-LAQQVYASDRANPEWTI 1388
 QY 1175 EHMETERLELAOKLNENYEVKSIITKERVKLKLOKSFETERDHLRGYIRIEATGLQTK 1234
 Db 1389 ARLETHLHGIQAKLOEDASYESL-----EAQNTLEQAR 1422
 QY 1235 EELKIAHIHLKHEQETIDELRRSVSEKTAQIINTODLEKSHSTKLOEIPVLHBEQELLPN 1294
 Db 1423 SA-----ALEBQAASQANQAAQKVQILEQKQREBEQKR-----QOQOQLQER 1471
 QY 1295 VKKVSETOETWN-ELELLTBSQSTTKDSTTLARIEMERLRLNEKFPQESQEEIKSTIKBRDN 1353
 Db 1472 FYELGQREQAQSQLELTSEA-----EESRQQLAGLRTYES 1509
 QY 1354 L-----KTKBEALVKHDQKHEHIRETLAKTIOESQSKOQOSLNMEK--DNETTKIVSEME 1407
 Db 1510 LLAHXSQLTATAQAREQMSHSSQELAEELRQQLDVEADLHRQVYDAKAALAKATELD 1569
 QY 1408 QFK-PKDSALIRIEMI GLSKRLOESHEDEKMSVAKEDDLQRLQOEVLQESDOLKENIK 1466
 Db 1570 ELECDLNSHVERAAATRELCQOLERSOE--LVAQTEELQRLNEEFQEVERESTLSR 1626
 QY 1467 EIVAKHLETEELVAHCCLEKEQETINELRVNLSEKETEISTITQKQLEAI-----ND 1519
 Db 1627 EVTL-----LKLQHDSEADQVLELQELRQMAQDKTEMNDLRTQIDALCANHSQELQ 1678
 QY 1520 KLQNKIOBI-----YEKEQOLNTKQISEVOENVNELKQFKEHAKAKSALQS 1566
 Db 1679 ALQRIAEALDTLQGNQTDQVVIETENKRLAEQSELQALQARQOHOQOQHHPAVOS 1738
 QY 1567 IE-----SKMELTNRLQESQEIQIMIKEEMKRVQEAQLQTERDQKENTKEI 1616
 Db 1739 QOHPPPSILFFGDAALAAPSPFDETAQPLRVSSLSAASAPPPISPPPTIE--DLQNVSDL 1796
 QY 1617 VAKKESQEKVQFLKMTAVNETQKMCIEIHLKEQFETQKLNLENIENTIRLTQILHE 1676
 Db 1797 ---EKHAQDLE---TKLLARN-----QNLAEQ-EERRLOEQRLSEVERLLSERTQ 1840
 QY 1677 NLEEMRSVTKERDDLRSVEETLKVER-----DQKENDRETITRLE-----1718
 Db 1841 QLADIQTANEERDLAALAEKLIQPAAPPTLDMFFGQAEETVPDASHHLDLGLPQTEPV 1900
 QY 1719 -----KQEBELKVHMLKEH-QETIDKLRGIVSEKTNIEISN--WOKDLEHSDALKAQ 1768
 Db 1901 VEPLIQPKAYLCOPKQEQIETQACTID--WGVDEDPWASAANEAPQDTVEHLH-----1952
 QY 1769 DLKIQEBELRIAHMLK---EQOETIDKLRGIVSEKTDKLSNMQKDLNSNAKLQEKIQE 1824
 Db 1953 -----TRIAQLEQLSNABQOKT-----ELQTKAAKLMKRLKE 1985
 QY 1825 LK-----ANBHQILITLKKDVNETQKKVSEMQKQKQIKDQSLTSLKLE-----IENLNL 1873
 Db 1986 YKTKATTATPTVTDNDLST-----IIEELKHQLOQESRLSKAEISQOHALEKEKL 2040
 QY 1874 AOL-----HENLEMK-----SYMKERDNLRRVETLKLERDQ-----LKESIQET 1915
 Db 2041 AKRIDVLTAGNDRMAEMKERQDMQVQYQARIRELOEKLS-QLDQWGEPAATVSSSLDGD 2099

Query Match 7.2%; Score 956.5; DB 1; Length 2238;
 Best Local Similarity 20.9%; Pred. No. 1e-18;
 Matches 515; Conservative 433; Mismatches 867; Indels 589; Gaps 99;

QY 166 DLTEVVYVSEWALK---WITGEEKS-RHYGET-----KMNQSSRSRHIFRM-ILES-- 213
 DB 48 DVTNRENASTQATKPDGVSQDSSPSQSGDTQTFAKQLQLRVPVSMESLPRSPIKESLF 107
 QY 214 REKGEPSNCEGVS---HLNVLVDLAGSERRAQTGAAGVRLKGCNINRSLFILGOVIK 270
 DB 108 RSKSEP-----LVRTSRESINQDLDCSAAFPDPSMESEADAPWN----- 151
 QY 271 KLSGQVGGFINVRDSKLTILQNSLGNPKTRIICTITPVSFDETLTALQFASTAKYMK 330
 DB 152 --SDGLSREQLQLRLRMERSLSYRG-----KYSLVTAFTQLQREK--K 193
 QY 331 NTPVYNEVSTDEALLKRYRKEIMDLKQLEVSLETRTAQAMEKDQALQLEBKD----- 384
 DB 194 KLGILLSQSDKSL---RRISEL---BELOMDQQAQKHLODFDACLXKDXQYISVL 245
 QY 385 -----LLQKVON-----EKIENLRMLVTSSSLTLOQELK 414
 DB 246 QTVSLKQRLQKLONGPMNVADPKPLPQELQAEVHGDTKEKMGVGPVGGTSAKTLEMLQ 305
 QY 415 AKRRVTVCLGKINKMKNKNYADQFNIPNTITTKHLSINLRIDEISVCSSESVFNS 474
 DB 306 QRVKQRE-----NLLORCKE-----TIGSHKEQCALLSEKALQELDERLQ 348
 QY 475 TLDLSEIWNPAKTLNQ-----ENISE-----INSRLADYNLV--- 511
 DB 349 ELEKMKELHMAEKTULTQLRDAKNLILBQBDQGMVITETKQMLTLEKEDEIAQLR 408
 QY 512 LDYQRLTEKEMELKUKENDLDEFALER-----TKKQDQEMQLIHEI 556
 DB 409 SHIKMTTQGBELR-EQKESERAPELEKALSTAQKTEPAQRMKMEDEQMKAVERA 467
 QY 557 SNLKNLVHRYVNDLENLSSKVELLREKEDQIKLQYIDSQKLNKMDLSYSLES 616
 DB 468 SEERLRLOHEL---SRVQEAASMAK---KNSBEQVAALQK-LHABELASKEQELSRLEA 522
 QY 617 IEDPKQMKOTLFDATVALDAKRESAFILR-----SENLEKEMKELATYKQWENDI 669
 DB 523 RE-----RELQEQMIALE-KSRSEYKLITOEKQESLAELELOKAILTESENKL 575
 QY 670 QLYQSQLEAKKQMDVLEKLSAFNEITKLTSLIDGKVPKDLNCLNLE-----GKITD 724
 DB 576 QELGQEAAYRTRILETSLSEKSLQ-SKTQS-----EHLAVHLEAKKNKHELTA 627
 QY 725 LQKELNKEVEENEARLEVILLSELKSLPSE-----VERLRKEIQDKSELHIITSEKDKL 780
 DB 628 LAEQHRTVEGLOQ-QODSLWTERLQSLSQHQAAVEELREKYQOEKDAL---LKEKESL 683
 QY 781 FSEVVHKSERVQGLLEIIGTKDDIATQSN-----YKSTDQBFQNFKTLHMD---FEQKY 833
 DB 684 FQ-AHIQDMNEKTLEKDKKQMELESVSELSEALRARDQALABELSVLRGADKMKQAL 741
 QY 834 KAVILENERMNGEIVNLSKAKFD--SSLGALKTELSYKTOBLOKQKTEVEORLNEMEQ 891
 DB 742 EAELEQRHOREVSGISSEQQSLTVRAEKALKDBLSRLGALLDERDEHLRERQARVD 801
 QY 892 LKQLENRDSPLQTVBEREKTILTEKLQOTLEEVKTLITQEKDKLQQLQESLQIERDQKSD 951
 DB 802 LEHLQKS-----AGELOALAKLDDLHSEQSAAREQAGAYFEOLAQMOK 847
 QY 952 IHDVTNNITDQOLNLAESLKHQHETINTLKSISSEESVSNLHMEENTGETKDFEFOK 1011
 DB 848 V-----LDLETEKSL-----LTKQVVMETHKXHVCEEELDAQRAVQQLERORSELEEK 896
 QY 1012 MVGIDKQDLKANTQTLTADVKNDNEIEOQRIFSLIOEKNELQOQMLLESVIAEKQLKT 1071
 DB 897 VRSIAQODSOLKNS-TVEKEAQRSQSMENKILQMRREQAQKEIILKTLTKSKEBSIS 955
 QY 1072 DLKENIEMTTEQBELRLGLDELKKQOEIVAQEKNHAIKKEGELSRCTDLRAEVEEKLKE 1131

DB 956 ILHEEYETFKNQEK-----RM-----EKIKQ 977
 QY 1132 KSQQLQEQKQOQLLNVOEEMSEMQKKNELNKNELKELTLEHMETERLELAQ----- 1186
 DB 978 KAKEMQETKKLLDQBAKLK-----ELENTVLELSQKE---KQNAQILEWAQANSAG 1028
 QY 1187 -----KLINENY-EEVKSIT-KERKVKLEKQSFETERDHLRGYIRIEIATGLOTKEE-- 1236
 DB 1029 ISDVTSLRSEENORQOIESUTGAHOKKLDVIEAWEKLSQAAELDKHAEQMEKEQGL 1088
 QY 1237 -----LKTARIHLKEHOETIDELRSVS-----EKTAQIINTQDLEKSH 1276
 DB 1089 GELRQKVRIVQSEKEELTKVEARLKEAVSQDVALAGLOQLKQKSAVIVS---LSRES 1145
 QY 1277 KLQEEIPVLHBEQELLPNVKVSETQETMNELELLTEQSTTKDSTTLTARTEMERLRNEK 1336
 DB 1146 QLSQVQEKL--EADLGCSLEKLSQEFELAKLADKLSQVSELSTGQVQAAMEKELQSC 1203
 QY 1337 FQESQEEIKSLTKERDNKTIKEALEVHKDQKHEIRETL-AKIQE---SQSQEQOSLN 1391
 DB 1204 KSLHELKSKLEDSKLSNLSLLELASQDSRCERTKALLEAKTINELVCTSRDKADAILA 1263
 QY 1392 MKEKDNETTKIVSE-----MEQFKPKDSALLRIEIMGL-----SKLOESHDWMS 1439
 DB 1264 RLSCQORHTATVGEALLRRMGQVSELBAQLTQLTQTEORTLKSSFQVNTQLEKEKOLKT 1323
 QY 1440 VAKKDDQLRQOEVLQSESQOLKE--NIKEIVAKHLETEEBELKVAHCLCKEQEETINELR 1497
 DB 1324 MKADIEGLITKEALQOEGGQROQAASEKSCITQLKKE-----LAENINAVTLR 1374
 QY 1498 VNLSEKETETSTIOKLEATNDKQNKIOEIVKEEQNLKQISEVOENVNELKQFKBEHR 1557
 DB 1375 EELSEKSEITASLSKQSLDGAQLESLSLSP-----SDKAEAISALS--KQH- 1418
 QY 1558 KAKDSALQSIBSKMELTNLROESQEEIQLIMIKK-----EEMKQVQAL 1602
 DB 1419 -----BEQELQALQALQELSLKVDALSKEKMSALEQVDHWSNKFSEMKKQASRL 1468
 QY 1603 QIERDQAKNTKEIVAKKESQEKYQFLKMTAVNETQKCEIHLKEQEFQTKNLLEN 1662
 DB 1469 AQHOSITDKLAQDVKATDAREKEBEQCLLKEDLDQNK--KPECLKEGEVRSKMEK 1526
 QY 1663 -----IETENRILTOILHENLEEMSVTKERDDLSRVEETLKVVERDQKKNIRETIT 1714
 DB 1527 KECDLTALKQTAARVE-----LED--CVTQKKEVESLNETLK-NYNQORDTEHSGIV 1578
 QY 1715 RDLEKQEEEL-KIVHMLKHEOFTIDKLRGIVSEKTNELSNMOKDLEHNDALKAQD--LK 1771
 DB 1579 QRLQHLBELGEEKONKVRABETVLRLREHVSLEAEILGTVKKELEHVNSSVKSRLDELK 1638
 QY 1772 -IQEELRI---AHMHLKQOQETIDKLRGIVSEKTKLSNMOKDLEHNSAKLQEKIQE-LK 1826
 DB 1639 ALEDKLELSAAKVELKRAE-----OKIAAIRKQL---LSQMEKTOYAK 1682
 QY 1827 ANEHOLIITLKKDVNETQKVSMEQIKKQIKD-----QSLTSLKLEINLNL 1873
 DB 1683 DTENRLELSAQIKREKQVHSLQEDKLNLESSPHEVPAVSKSMQSVAAASP-EQAPDS 1741
 QY 1874 AQELHENLEEMKSVMKERDN---LRRVEE-----TLKLERDQKESQETKARDL 1920
 DB 1742 QDTHKACKERLCMLQRRLLSEKEKLRLRLEQGEARPSQPEAQHALLSGKLDCTARQL 1801
 QY 1921 -----ELQOELKTARMLSKHEKE-----TVDKLREKISEKTI--- 1952
 DB 1802 EDHVLGICLPEELEKXMKCSLIVSQPMGEETGNNTGVKNWASVVDVSKTLQEKELTCQ 1861
 QY 1953 -----QISDIOKDLKSKDELQKKIQBELQK---ELQLLRVKED-----VNM 1991
 DB 1862 ALBQRVKELSLDLVRGAHRLEVEKILTLYEKSQSQOQMDGENCKCVELDERPEENSQ 1921
 QY 1992 SHK---KINMEQOLKQFEPNLYCKBMDNFQITKHLSELEIRIVAKE-RDELRIKE 2047

Db 1922 SHZQSVGVTDGLRDLB-SKLTGAERDKQKUSKEVARLQKELRALRREHQBELDKR 1980

Qy 2048 SLKWRDQFIATLREMIARDQNHQVPE---KLLSDGQOHLMESLREKCSRIKELKR 2104

Db 1981 ECEQEAR-----EKLQSQEDLELKHSTLQLMREFNTQLAQKEQLERTVQETIDK 2033

Qy 2105 YSE-----MDDHYECINRLSLDI-EKEIEFHRIMKLLKYVLVYVTKIKESQHCINKFEM 2158

Db 2034 AQEVEAELSHQEQETQQLHRKIAEKEDDLRTARRVEEILD----- 2075

Qy 2159 DFIDEVEKOKELIKIQLHQQDCVPSRELRLDLKLNQMDLHIEILKDFSESEFPISKT 2218

Db 2076 -----ARREMTGKVTDLQ-----TQLEELQKKYQORLEQESTKD--SVTILELOT 2120

Qy 2219 EFQQ-----VLSNRKEMTQFLEEMLNFTRDIEKLNQKQKENDRICQV-----NNFFNN 2267

Db 2121 QLAQKTTLISDSKLKEQLRE-----QVNHLEDRKLKRYEKNAACATVGTYPYKGNLYHT 2174

Qy 2268 RIATINNESTEF-----ERSATISKWBQDLKSLKEKNEKLFKNYQT----- 2310

Db 2175 E-VSLFGEPTFEYLRKVMFEYMMGRETMTAKVITTVLKFPDQQAQKILEREDARLMSW 2233

Qy 2311 LKTS 2314

Db 2234 LKTS 2237

RESULT 9

RBP2 PLAVB

AC Q00799; Q9NZM3; PRT; 2867 AA.

DT 01-APR-1993 (Rel. 25, Created)

DT 15-MAR-2004 (Rel. 43, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Reticulocyte binding protein 2 precursor (PVRBP-2).

GN RBP-2 OR RBP2.

OS Plasmodium vivax (strain Belem).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.

OX NCBI_TaxID=31273;

RN [1]

RP SEQUENCE FROM N.A., AND REVISIONS TO 2438-2439.

RX MEDLINE=20299192; PubMed=10838229;

RA Galinski M.R., Xu M., Barnwell J.W.;

RT "Plasmodium vivax reticulocyte binding protein-2 (PVRBP-2) shares structural features with PVRBP-1 and the Plasmodium yoelii 235 kDa rhoptry protein family.";

RT Mol. Biochem. Parasitol. 108:257-262(2000).

RN [2]

RP SEQUENCE OF 1189-2439 FROM N.A.

RX MEDLINE=92315338; PubMed=1617731;

RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;

RT "A reticulocyte-binding protein complex of Plasmodium vivax merozoites.";

RT Cell 69:1213-1226(1992).

CC -!- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to human reticulocyte cells.

CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL; AF184623; AAF76525.1; -

DR HSPF; P03069; IGCM.

KW Malaria; Receptor; Signal; Transmembrane; Repeat.

FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 2867 RETICULOCYTE BINDING PROTEIN 2.

FT DOMAIN 22 2805 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 2806 2826 POTENTIAL.

FT DOMAIN 2827 2867 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 44 133 ASN-RICH.

FT DOMAIN 560 758 LYS-RICH.

FT DOMAIN 1112 1285 LYS-RICH.

FT DOMAIN 2758 2785 7 X 4 AA TANDEM REPEATS OF H-D-D-T.

FT REPEAT 2758 2761 1.

FT REPEAT 2762 2765 2.

FT REPEAT 2766 2769 3.

FT REPEAT 2770 2773 4.

FT REPEAT 2774 2777 5.

FT REPEAT 2778 2781 6.

FT REPEAT 2782 2785 7.

SQ SEQUENCE 2867 AA; 331433 MW; 6E7D8CA71AFBFD3 CRC64;

Query Match 7.2%; Score 956.5; DB 1; Length 2867;

Best local similarity 21.2%; Pred. No. 1.2e-18;

Matches 567; Conservative 486; Mismatches 932; Indels 693; Gaps 119;

Qy 329 MKNTPVYNEVST--DEALLKRYRKEIMDLKKQLEEVSLERAQAMEKDLQAL----- 379

Db 248 LENPQYNDISKYDEK-VKEYKKIEDMQICLDSYIRKFAIMKANLGNALNGIYI 306

Qy 380 -----LEEKDLQKVQNEKIEINLTMLVTSSSLTL-QQELKAK--- 416

Db 307 HWWYLTCSKTYDDIVKEVAIINDEDEKKSIFMDNMKK--IHKSAIDTLKKQKAEKLN 364

Qy 417 --RKRVTWCLGKINQM-----KNSNYAQDF--NPT-----NITTKT----- 450

Db 365 SLDSKRTETIIGEHMIEKFNHLTKIRYASAFATKSIPLQKVESDIYRVELTKLPYVA 424

Qy 451 -----HKLINLREIDSVCSDESDFNTLTLSEIENWPAKTLNQINSEINSLRA 505

Db 425 KHYADPKFSLEHLKMPENLSKSKMLYSTFFKLEGDLNKINTLMGSEQSDLTSLIA 484

Qy 506 DYDNLVDLYEQL-RTEKEEMELKKEKNDLDEFEAL-ERKTKKQDQEWLIHISLNK-- 561

Db 485 DSEKIIKSAESLINSSEI-----AKVALDSNEKINEIKKNYDQNIILKVRFINKSNGL 539

Qy 562 --LVKRVYNOQLENNELSSKV-ELLREKEDQIKKLOEYIDSKLENTKMDLSYLESTE 618

Db 540 ITSVMGTSQLSQSDQKQETKIEIKKKKKDLERGERKEFINMN-EIKKKKSNSSNST 598

Qy 619 DPQMKQKTLFDAETVALDAKESAFLRSENLEKEMKELATYKQMDNIQYQSULEA 678

Db 599 NSEFTDKLELETFEGNLTVMKGYLOEIEDIKVKNEDRSKQIIEOHLK-VTSDNRD 657

Qy 679 KKKQVLDLEKELQSAFNEITKLTSLIDGKVPKDLLCNLEEGKITDLOKELNKEVEENEA 738

Db 658 NVKTLISKNDIEIKYIEKIEKLN--DAPSKD-----KFTTEKTNLQNKVKKIIDEFH 709

Qy 739 LRREV-ILLSELKSLPSEVERLRKEIQ--DKSEELHIITSEK-----DK 779

Db 710 -KEDLQLLNLSLKFYEEHQKLYNEASTIEIKDLHQTKKEYEKLERKMFKNFQILDK 768

Qy 780 LFSVWHKESRVQGLLE-----IGTK-----DGLATQSNYKSTDDQEPNPKT 824

Db 769 LNTEDNLNLEKNIYBEQTNVINKVMSDSLTNLTAEVDNLSALDGYRADETELTKYN 828

Qy 825 LHMDFEQKYKMWLEENE-----REKLTITEKLOQTLREVKTLT-QEKDLDK 852

Db 829 RINERKEKFLSTLKEQEDDIPDGKNIYEEYNNHKNVWVKNKHKISSINQCNENIIEAK 888

Qy 853 EAQKFDSSLGALKTELSTYKTELQEQ-----KTREVERLERLMEQ-----LKEQL 896

Db 889 NLETFTNLVQTLDAHTGKQKQKVHDLQKFTNLKLNLELESFGKSLNGSASTYTKQI 948

Qy 897 ENRDSPLQTVB-----REKLTITEKLOQTLREVKTLT-QEKDLDK 935

Db 949 ENIRKNIDTTIKSLNFAKNSSSKLSLENIKNKADLIKKLDQHTQIEKHTFENEEMS 1008

Qy 936 QLOESLIQIERDQLKSDIHDVTWNVIDTQELRNLAKESLKHQETINTLKSKEEVSRL 995

Db 1009 PLLSVIKKKNRVESDMSE-----ELIKQNTKINAILLEYNKKDRFN 1052

```

1050 996 HMEENTGETYDFPOORWVGIDKKODLEAKNTOTLTADVKNELI-BQQRKIFSL-----IQ
1053 1053 GDDTELEDDDPKQCO--DAQOEIKKLTNTNVLDNGINVIKQEHKVIILSENHIT 1110
1051 1051 EKN-----ELQOMLESVIAEKEQL-----KTDLKENIEMTIOEERLRLGLDELKQOEI 1100
1111 1111 EKKKLINEKIQNVNSINEMTKLGLLKINEDIKNSRDTTIKSK-----IQEPEKKVQTI 1165
1101 1101 -----VAQEKNAHAIKEGELSR--TCDLRAVEBEKLKESQOLQKQOOLLNVQOEMSEM- 1153
1166 1166 FGSIDVANKKIDAIKKEHDVNDKDEKVKOTSFDEKKSISIKAYEKMGMNTLKELEKMD 1225
1154 1154 -----QKLINEIB-----NLKN-ELKNKELTLEHME-----TERLELAQKLINE-- 1190
1226 1226 DEKNIEKEVEEAGIQVRIIFIDHVNLMNDEVEKSKIVMEKIELYKKEIDEIKQKTNEYK 1285
1191 1191 -----NY-----EYKVSITKERVKJEL-----QKSFETERDHLRGVIRIEATGLQTKE 1235
1286 1286 QGDTSNFYTYEQNSATQSKAKIEQFINIATTKGTSDTSQD-----INLESIKERVHK 1340
1236 1236 ELKIAHILKHEOBTIDELRSV-----SEKTAQII--NTQDL-----EKSHTKL 1278
1341 1341 NLQI-----VKQESNSMEEMRKQILSMKDLILNNSSETIAKEISNNTQNALGFRENATIK- 1395
1279 1279 QBEIPVLHESQELLPNV-KKVSETQETMNELELLELTQOSTTKDSTTLARIEMERLRLNEKF 1337
1396 1396 -----LNKTDLLQORVAMIEEAKAHKNIDIALEDA--QIDTEVSKIE-----QINREI 1443
1338 1338 QESOEIKS-----LTKERDNLTKTKEALE-----VKHQDLKEHIRETLAKTOES 1382
1444 1444 MNKDEIKSYLSIBEIKYKDKCTTBEISNKRKDKIEFLEKPKPEESNSKNVINEINEN 1503
1383 1383 QSKQEQSINMKEKONETKIVSEBQPKDQKDSALLRI--ELEMGLSKRLOESHEDEKSV 1440
1504 1504 IRNSEQVIL--KDIEAEQAQSTKVELPHKHETTISNIPKSEIILGVETKSOKKINKAEDI 1561
1441 1441 AKEKDDLQRLQVLESQDLKENIKELIVAKHL--ETBEEL-----KVAHCCLKEQE 1490
1562 1562 MKE-----IERHSEIQTKVGFQENLNKLNBNPHYNDNAEDELNDKSTNAKVL-----IE 1612
1491 1491 ETINELRVNLSEKETEISTIQKLEAFINDKLNQKIQEIEYKEEQNIIQI SEVQEN----- 1546
1613 1613 TNLESVKHLS-----EITNIQGGGEKIYSKAKINDIMQIKATSENTAETKLEKVDQDSNY 1668
1547 1547 VNELQKQFKEHRK--AKDSALOSIESKMLELNLRLQESQEEIQTIMKEK-REM-KRVOEA 1601
1669 1669 VNYLNQITITERNLIVTEKNRLNGIDSTITNEGALKESKGNVEIGFLEKLEIGKVRKLK 1728
1602 1602 LQIERDQKENTKEIVAKMKEQKEKEYOFLKMTAVNETQEKMCET-----BHL 1649
1729 1729 VDIITKKSINSTGVNFSFLNFDLQYDYNK--NINDYENKMGELIYNFEGLSLNKISENL 1786
1650 1650 KQOFET-----QKLENIEETE-NIRLTQILHENLEMRSVTKERDRLR 1692
1787 1787 RNASENTSDYNSAKTLRLAEQKEKVNLLNKBEEANKYLRDV--KKVSEFRFIENKBSLD 1844
1693 1693 SVEETLKVR-----DQKLENLRETIITR-----DLEKO-----BEL-KIVMHLKE 1732
1845 1845 KINEMIKKEQLTVNEGHGNVQKLVENIKELVDENNISDLKQATGKNESIQKITHSTLKN 1904
1733 1733 HOSITDKLRGIVSEKTEINISNMQKOLEHNSDALKAQDLKIQBELRIAHM-----HLKEQOE 1788
1905 1905 KAKTI-----LGHVDTSAKYVGIKITPELALTELLGDAKLKTAQE 1944
1789 1789 TIDKLRGIVSEKTKLSNMQKOLE-NSNAKLQKQKIQ-ELKANEHQILITLKKV----- 1839
1945 1945 LKFESKNVNVLETNNSKNTNELDVKHNIQDAYKVALEITLAHSDEIDTKQKSKLIEMG 2004
1840 1840 NETQKQVSEWELKQKQIKQDQSLTSLKLEIENLNL--AQELHENLEEMKSVKMERDNLRRV 1897
2005 2005 NQIYLVKVLINQYRNKI---SSIKSKEEAVSVKIGNVSKHSELSKITCSKSYDNIAL 2061

1998 1998 BETKLERDOLKESLOETKAR-----DLE-IQOELKT-----ARMLSKHEKE 1938
2062 2062 EKOTELQ--NLNRSPTQKTNNTNSDKLEKIKITDPSLKNAUKTLTGEVNAKASSDNHE 2119
1939 1939 TYDKLREKISEKTIQISDIQKDLK---SKDELOKQIKLOKQKELQLLR--VKEDVNMNH 1993
2120 2120 HVQSKSEPNPALSEKEBETDIDSINTALDELLKKGRTCVSRVKLIKIDVTVKESDDT 2179
1994 1994 KKNEMEQ-----LKKQF-----EPNVLCCKCEMDNFQLTQ 2023
2180 2180 ELINTIEKNVAVLAIKKNYEDTVQDVLTLNBEHFNTKQVSNHEPNFNDKSNKSSBELTK 2239
2024 2024 KLHES-----LEEIRIVAKERDELRIKESLQMRDQFIATILREMIARDRQNHQVPEK 2077
2240 2240 AVTDSKTIISKLKAGVILEVENETEMTIERSSAK-ELEALYNELKNKKTSLNEIYQTSNEV 2298
2078 2078 RL-----LSDGQ-----OHLMESLREKC-SRIKELLKRYSEMDD 2110
2299 2299 KLOEMKSNADKYIDVSKIFNTVLDTKNSIVTNQHSINNVDKLGKQLQELIDADS--- 2354
2111 2111 HYECLARLSLDLEKEIFPHRIMKKLVLSVYVTKIK-EQHECINCKEMDFIDEVEKOKE 2169
2355 2355 -----SFTLESIKKFNIEYSHIK-----TNIGELOLQQTNKSEH---DNVAKHKE 2397
2170 2170 LLIKIQHL-----QQCDVPSRELRLDLKLNQM-----DLHIEELIKOF 2208
2398 2398 ---KIVHLINRVESLKGDKVKNHDDQYMKKLNASLNDINKNTNSINISDEELKLLKKV 2454
2209 2209 SES-----FPSIKTEFOQVLSNRKEMTQFLEEWLN 2239
2455 2455 EENDQCKNNTQNFISIMKRVEDLNRRFTENLPE-KEKLIQIENNYEISSIFSE-IN 2512
2240 2240 TRPDIEKLKNGIQKEND-RICQVNNF-----FNNRIAIIMNESTEFEBRS 2283
2513 2513 LQ-DVDEFVAKIKHQIDAERKASVNNVREAEKIRTAIONVTSYDTIELISRLSEMNNVLERI 2571
2284 2284 ATISKWEQDLKSLKEKNEKLFKNYQTLKTSLASGAOVNPTQDNKNPHVTSRATQLTPE 2343
2572 2572 TTRTKMDQLKLSLSPN-----TSL-----NLNARTHVRKSEDIK--QLNSH 2613
2344 2344 --KIRELENSLHEAKESAMHESKIIIMQKQLEVTNDIIAKLQAKVHESN-KCLEK---T 2397
2614 2614 IGEITELNTYAHEV--MTYLENELNKLKQLEIER---AKLETSTPSGMKAKEEKVPP 2667
2398 2398 KETIQVLQDKVA-----LGAKPYKEBEDLKMVKIDLEKMKNAKEFEKEI 2444
2668 2668 KETENRAQDNLASVPQKLTEDNTQOMPENSVDNLASVPQKLTEDNTHQMPENRVQEDSI 2727
2445 2445 SATKATVYQKEVIRILLRENLRRSQQAQUTSVISEHTD 2482
2728 2728 SAPQEQVEY-----STLAVPENDETTEESEHDD 2756

RESULT 10
KF4A HUMAN
ID KF4A HUMAN STANDARD; PRT; 1232 AA.
AC O95239; Q9NNY6; Q9N24; Q9UMW3;
DT 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chromosome-associated kinesin KIF4A (Chromokinesin).
GN KIF4A OR KIF4
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymphocytes;
RA Villard L.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
```


Db 909 FAEIETELQAEVMEQ-----QHKEVLYLLSQLOOSQMAE----- 945

QY 935 KQLOESLOIERDOLKSDIHDVTNNMIDTQBLRANALLESIKQHOTINTLKSIS--EVS 992

Db 946 KQLESVSSEKEQQLS-----TLKQDEELEKREVCENQOQLRENEIIKQKULLQVAS 1001

QY 993 RNLHMENT-----GETKDFEQKQWGI-DKQDLEAKNTQTLTADVK 1034

Db 1002 RQKHLPKDTLLSPDSFEYVPPKPSRVKEFKLEQSMIEDLKYSERHSVNEHEDGDG 1061

QY 1035 DNEIIEQORXIFSLIQ 1050

Db 1062 DDEGDDEWPTKLK 1077

RESULT 11

KF4A MOUSE

ID_KF4A_MOUSE STANDARD; PRT; 1231 AA.

AC P33174;1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 32, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Chromosome-associated kinesin KIF4A (Chromokinesin).

GN KIF4A OR KIF4 OR KNS4.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE

RP SPECIFICITY.

RC STRAIN=ICR; TISSUE=Brain;

RX MEDLINE=95014709; PubMed=7929562;

RA Sekine Y., Okada Y., Noda Y., Kondo S., Aizawa H., Takemura R.,

RA Hirokawa N.;

RT "A novel microtubule-based motor protein (KIF4) for organelle

RT transports, whose expression is regulated developmentally.";

RJ J. Cell Biol. 127:187-201(1994).

RN [2]

RP SEQUENCE OF 91-240 FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=93077686; PubMed=1447303;

RA Aizawa H., Sekine Y., Takemura R., Zhang Z., Nangaku M.,

RA Hirokawa N.;

RT "Kinesin family in murine central nervous system.";

CC -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR

CC SPINDLE STABILIZATION.

CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic

CC chromosomes.

CC -!- TISSUE SPECIFICITY: EXPRESSED IN PYRAMIDAL CELLS IN JUVENILE

CC HIPPOCAMPUS, GRANULAR CELLS IN JUVENILE CEREBELLAR CORTEX AND IN

CC ADULT SPLEEN.

CC -!- SIMILARITY: Belongs to the kinesin-like protein family.

CC Chromokinesin subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; D12646; BAA02167.1; -.

CC PIR; A54803; A54803.

CC HSP; P17119; 3KAR.

CC MGD; MGI:108389; Kif4.

CC InterPro; IPR001752; kinesin_motor.

CC Pfam; PF00225; kinesin; 1.

CC PRINTS; PR00380; KINESINHEAVY.

CC SMART; SM00129; KISc; 1.

DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.

KW Motor protein; Microtubule; ATP-binding; DNA-binding;

FT DOMAIN 1 350 KINESIN-MOTOR.

FT NUCLEAR 351 1000 COILED COIL (BY SIMILARITY).

FT DOMAIN 1001 1231 GLOBULAR.

FT NP_BIND 88 95 ATP (POTENTIAL).

FT CONFLICT 112 112 I -> S (IN REF. 2).

SQ SEQUENCE 1231 AA; 139551 MW; F34F2C2D21158FE4 CRC64;

Query Match 7.1%; Score 952.5; DB 1; Length 1231;

Best Local Similarity 26.0%; Pred. No. 7.5e-19;

Matches 367; Conservative 223; Mismatches 491; Indels 329; Gaps 51;

QY 7 VAVCVVRPLNSREESLG-ETAQVYKTDNNVLYQVDSKSFNFDRVFGHNETTKNYEE 65

Db 10 VRVALRCPLVSKIEKEGQTCLSFVPGEPQV--VGNDSKFTYDFVDFSTEGEEVNT 67

QY 66 IAAPIIDSAIQGYNGTIFAYGQTASGKTYTMGSG---EDH---LGVIPRAIHDFQIK 118

Db 68 AVAPLKGKGVFNATVLAIGQTSKYSWGAYTAEGHDSAGIVIPRVQLLPKEIN 127

QY 119 KFPDREFLLRVSYMEIYNETITDLCGT-QMKPLIREDVNNVYVADLTETEVVYSEM 177

Db 128 KKSDFEFTLVSYLEIYNEBILDLCSREKATQINIREDPKEGKIVGTEKTVLVASD 187

QY 178 ALKMITGKSRHYGETKMNQSRSHITFRMILESREKGEPSNCEGSKVSHNLVDLA 237

Db 188 TVSCLEQGNNSRTVASTAMNSQSSRSHTAFTISIEQRK---NDKSSSFPSKHLVDLA 243

QY 238 GSREAAQTGAAGVRLKGCNCNRSFILGOVKKLSGQGVGFNYRDSKLTRILQNSLG 297

Db 244 GSERQKTKAEGDRREGINRINRGCLGENVISALGDDKGNFVPRYDRLTRLLQSLG 303

QY 298 GNPRTIITCTTPV--SPDETTLQALPASTAKYMNTPYNEVSTDEALLKRYKEIMDL 355

Db 304 GNSHTLMACVSPADSNLEETLNLTVADRARKINKPIIN-----IDPQAAELNHL 355

QY 356 KQLEEVSL-----ETRAQAMEKDQLAQLLEKDLQKQVONEKIN 396

Db 356 KQVQQLQLLLOAHGGTLPGDINVPSENIQSLMERKQ---SLVEENKLSRGLSEAAQ 413

QY 397 LTRML---VTSSSLTQQLKAKRRRTVTCGLGINK---MKNSYADQFNIPNTYTK 449

Db 414 TAQMLERILTEQANEKNAKLELRHAACKVDLQKLVTLEDQELKENIICNLQOV 473

QY 450 THKLS-----INLREIDESVCS-----BSDFVNTLDTLSE-----LEWNP 486

Db 474 TAQLSDEAAACMTATIDTAGEADTQVQSSPDTSRSDVFS-TQHALRQMSKELIELNK 532

QY 487 A-----TKLNLQENISELSLRADYD---NLVDVYEQLRTEKEEMELKIK-EKNDL 534

Db 533 ALALKEALAKKNTQ--DNQLQPIQFOQDNITKNLESEVLQREKEELVLELQAKKDA 590

QY 535 DEFEALERTKDKDQEMOLIHETSNLKNLVKREVNQDLENELSSKVELLRKEQIKKL 594

Db 591 NQAKLSERRRKLQLE--GQIALDKKLQ-----EQSKLLKKESTEHTVSKL 637

QY 595 QBYIDSQKLENKMDISYSLIESIEDPKQKQTLFPAETVAL---DAKRESAFLRSE-NLE 650

Db 638 NQELRMKQNRVQL-MRQMKEDAEPKQWQKQ-KDEVQLKDERDKRQYELLKLRNFQ 695

QY 651 -----LKEQKELATTYKOMENDIQLYQSLQLEAKKQVQDLEKELQSAFNEITKLTSLID 705

Db 696 KQSNVLRRTTEEAANKKALQ-----KQEVAKKETQSRGNESTAARM---744

QY 706 GKVPKDLLCNLELEGTITDLQKELNKEVBENEALREEVILLSELKSLPSEVRLKEIQD 765

Db 745 -----KNWLGK-----EIVVWSTEEAKRHLNGLLEERKILAQDVAQL-KEKRE 787

QY 766 KSEELHITSEKDKLFSVVKESRVQGLLEELGKTKDDLATTSQSNYSKSTDQEFQFKTL 825

Db 788 SGENPPLKLRRTSYDEIHGQDS---GAEDSIK 819

Qy 826 HMFQYKQVLEENRNQEIIVNLSKQAKFDSSIGALKTELSTVKTQELQKRETVQER 885

Db 820 ---QIESLETELBSAQI---ADLQOK 841

Qy 886 LNEHEQLKEQLENDSPLOTVVERKTLITEKLOOTLREVKTLTOEKDDKLOQLSLOTER 945

Db 842 LLD-----AESDRPKQWESIATILEAKC-----AIKVLVGLVSSKILVSKLSSSL 889

Qy 946 DQLKSLDHTVNMNIDTQELRNALSLK-----OHOETINTLKSKEEVSERNLHM 997

Db 890 NQSKASIDVQKMLFEQNHPIAKTETELKEELVKVEQHQEKVLYLLSLOQSQOMTEKOL 949

Qy 998 EENTGETKDFQKQWGDIDKQDLKAKNTOTLTADVKNDIIEOQRIFSLIQEKNELQ 1057

Db 950 EESVSEK---EQLLSTLQCEELRKMQ-----EVCEQNOQ---LLENSAIKQ 993

Qy 1058 MLESV-IAEKEQLKTDLKENIEMTIENQ-----BELRLGDELKQKQ 1098

Db 994 KLTLQVASKQ--KPLTRNIFQSDSSPEYIPPKPCRIKEKCLEQSFVAGELQVYSE 1051

Qy 1099 EIVAQEKNAIKKEGELSRCTDLAEVEKLKX-----KSQQLQKQOQLLVQEE 1149

Db 1052 PSVAEQNEDSDHADREWTPTKLIVKSKSIQSCSKGCGKQCGCRKQKSDCNVSCS 1111

Qy 1150 MSEMOKKINEIENKLNKELTLEHMETERLELAQKLNVE-----EVKS----- 1197

Db 1112 CDPTK-----CRNRHQND-----NSDAELNQDSNSFKLEDPTEVTSGLSFFHP 1157

Qy 1198 --ITKERVKLKEI-----QKSFETERDHRGVIYREIATFGLQTKBELKTAHILKE 1246

Db 1158 ICATPSSKILKEMCDADQVLQKPMFVSSSDH-----PELKSIASESQE 1201

Qy 1247 HOETIDELRSVSEKTAQIINTQDL-EKSH 1275

Db 1202 NKAIGKKKRALASNTSFFSGCSPIQESH 1231

RESULT 12

KF4A XENLA

ID AC Q91784; Q9PS10; STANDARD; PRT; 1226 AA.

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Chromosome-associated kinesin KLP1 (Chromokinesin).

GN KLP1.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

OC Xenopodinae; Xenopus.

OX NCBI TaxID=8355;

RN [1]

RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.

RC TISSUE=Oocyte;

RX MEDLINE=95236444; PubMed=7720067;

RA Vernos I., Raats J., Hirano T., Heasman J., Karsenti E., Wylie C.;

RT "Xklp1, a chromosomal Xenopus kinesin-like protein essential for spindle organization and chromosome positioning."

RL Cell 81:117-127(1995).

RN [2]

RP SEQUENCE OF 9-338 FROM N.A.

RX MEDLINE=93246065; PubMed=8482413;

RA Vernos I., Heasman J., Wylie C.;

RT "Multiple kinesin-like transcripts in Xenopus oocytes."

RL Dev. Biol. 157:232-239(1993).

CC -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR SPINDLE STABILIZATION.

CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic chromosomes.

CC -!- TISSUE SPECIFICITY: Expressed in oocytes, eggs, testes and

CC brain.

CC -!- SIMILARITY: Belongs to the kinesin-like protein family.

CC Chromokinesin subfamily.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC ENBL; X82012; CAA57539.1; --

DR PIR; I51617; I51617.

DR HSP; P17119; 3KAR.

DR InterPro; IPR001752; kinesin_motor.

DR Pfam; PF00225; kinesin; 1.

DR PRINTS; PR00380; KINESINHEAVY.

DR SMART; SM00129; KISC; 1.

DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.

DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.

KW Motor protein; Microtubule; ATP-binding; DNA-binding;

KW Nuclear protein; Coiled coil.

FT DOMAIN 1 350 KINESIN-MOTOR.

FT DOMAIN 351 1006 COILED COIL (BY SIMILARITY).

FT DOMAIN 1007 1226 GLOBULAR.

FT NP_BIND 87 94 ATP (POTENTIAL).

FT CONFLICT 163 163 I -> L (IN REF. 2).

SQ SEQUENCE 1226 AA; 138923 MW; 7F0275FCF3316697 CRC64;

Query Match 7.0%; Score 935.5; DB 1; Length 1226;

Best Local Similarity 24.6%; Pred. No. 2.1e-18;

Matches 341; Conservative 261; Mismatches 478; Indels 309; Gaps 45;

Qy 3 EBG-AVAVCVRVPLNLSREELG-ETAOVYKTDNNVIYQVQDSKSFNDRVPHGNETTK 60

Db 4 DEGIPVRVALRCRPLVPKKNEGCKMCLTFVPGEOQVI--VGTEKSFYDYVDFPSAQE 61

Qy 61 NYVERIAPIIDSAIQGYNGTIFAYQGTASGKTYTMGSEDH-----LGVIPRAIHDI 113

Db 62 EYNSAVAPLIKGLFKGYNATVLAQGSGKTYSGGAYTHQENEPVGVIPRTVIAL 121

Qy 114 FQIKKFPDRELLRVSYMEIYNEITDLL-CGTQMKPLIREDVNRNVYVADITEEVV 172

Db 122 FREIHQPEWEPNKLVSYLEIYNEIBDLVLAARDKNTTISIREDPKEGKICGLTERDV 181

Qy 173 YTSEMAKWTGKEKSRHYGETKMNQSRSHITFRMILESEKGEPSNCEGSKVSHLN 232

Db 182 KTALDTLSCLQGNSSRTVASTAMNSQSRSHAIFTISIQKREGDKNN---SFR-SKLH 237

Qy 233 LVDLAGSRAAQTGAAGVRLKEGCMNINSLFILGOVIKKLSD-GQVGGFINYRDSKLTRI 291

Db 238 LVDLAGSERQKTKAEGDRLEKISINRGLLCLGNVISALGDESCKGKGFVPYRDSKLTEL 297

Qy 292 LQNSLGGNPKTRIITITPV--SPDETITAFQASTAKYMQNTPYVNEVSTDEALLKRYR 349

Db 298 LQDSLGGNSHTIMTACVSPADSNMEETLNTLYRADRKIKNKPTVN-TDQAAELQRLK 356

Qy 350 KEIMDLKKQLEEV-----SLETRAQAMEKD-----QLAQ 378

Db 357 LOVQELQVLLLOAHGGTLPVLNSMPSNLOSMLERNKLEKNGKUSRELGEAAVQTAQ 416

Qy 379 LLEKDLQKQVNEKI-----ENLFRMLVTSSTLTQ-----QELKA 415

Db 417 FL-EKIIMTEQONEKLGSKMEELQHAACKVNLQRLVETLEDQELKDNVEIQNLQVIV 475

Qy 416 KXKERVTCGLKXNKMNKNVADQFNIP-----TWITTKHKLISNLLREI 461

Db 476 QLODESSGIAGSIAMDEE--AASFVPVEEDSGEKSSDGTFTNHALRQAQLSKELI-EL 532

Qy 462 DESVCSSESDVFSNTLDTLSIEWNPAKLNQENIESELSRLADYDNLVLDYQLKRYR 521

Db 533 NKALVMEALAKWQAQNDRQLEPIQSEYLNINIKHLESEVGLQKEBELIALHSAKDN 592

QY 522 EEWELKKEKNDLDEFEALERTKQDQEMQLIHEISNLKLVKRE-----VYNQDLE 574
 Db 593 NQAKLSERRRKLQLEGGQWTELUK-----KLGESKLLRLRSTETVAKMKEIQ 644
 QY 575 NELSSVLLREKEDQIKLQYIDISQKLNIRKDLSSLESIEDPKMKQTLFDATVA 634
 Db 645 GMDQVQLMRQKEDAEKFRTWKQKTEVIQI-----KEKDRKQVYELL----- 690
 QY 635 LDARSAFRLSRNLKELKQKELATYKQMDNDIQLYQSLKAKKQVQDLEKELQSAF 694
 Db 691 ---KLERDFQKQVQV-LRRKTEEAASANKELQKQKAEKQKQDSQ---SKGMEGAA 743
 QY 695 NEITKLSLIDGKVPKDLLCN-LELECKITDLOKELNKEVEENALREEVILLSELKSLP 753
 Db 744 SRV-----KNWLANEVELVSTEAQRLNDLEDRKILAQDI---AQLKQKT 788
 QY 754 SEVERLRKEIQDKSEELHIITSEKDKLFSEVHKESRVQGLLEIGTKDKDLATQSNYK 813
 Db 789 DAGERIPTKIRRT-----YTVAELENLEEEASVTK----- 819
 QY 814 STDQEFQNFYKTLHMDFEQKQVWLEENRQWQEIWNLSKEAQKEDSSIGALKTKELSYKTQ 873
 Db 820 -----QESLETEWELRSA 833
 QY 874 ELQKTRVORRLNEMQLEKLENRDSPIQTVREKTLITEKLIQOTLEEVK---TLTQE 930
 Db 834 QIADLQKLLDADGEEWVKRWETISNIMEAKALKYLITELVSSKVGSKLESSVKQN 893
 QY 931 KDDLKQLESQIERDQLKSDIHDVTVMNHDITQBLRNALSLKQHQETINTILSKSISE- 989
 Db 894 RAHVADLQKNIFEERNQW-----AEMETEHSQSL---MQLEHQHKEKILYLLSQLOK 943
 QY 990 EVSNRLNMEENTGETKDFQKQV-----GIDKKQDLKAKNTQTLTADVKQNEILEQ 1041
 Db 944 QASVPVTIELPABEITEREQMLERLKFQDEIEKMKALCEKQKQLL---QENDMTYK 999
 QY 1042 QKTLFSLQKNEQLQOMLESVIAEKEQLKTD-----LKENIEMTIENQEBLR 1088
 Db 1000 KLALHVASGK-----LHNILPAAEICSPPSPDFIPPKRGKRTWAKSAVILEDL- 1053
 QY 1089 LLGDELKQKQEIQAQKQKHAIKKEGELSRTCORLAEEVEKL---KKSQOLQEKQOQLLNV 1146
 Db 1054 LSESESEESDDKQWEPGNQSKSLTKSCCKARCGNKGCRKTKQ-----NC 1104
 QY 1147 QEEMEMQKQKNEINLKNELKNKELTLEHMETERLELAQKLNENYEVKS----- 1197
 Db 1105 SDDCFCDPSCNRNDRNDHDEGKHQDOSLE-SENSKID-----YPDVTAGGSFTTTPC 1155
 QY 1198 ITKERKVLKQKSPETERDLRGVIREATEATGLQTKBELKLAHILKHEQHTIDELRRS 1257
 Db 1156 VPTTKVLKREISD-----IGQVLSIKLQKRPSTASASVMSQEN---QTS 1199
 QY 1258 VSEKTAQII 1266
 Db 1200 ILTKKKKVL 1208

RESULT 13
 KF4A_CHICK STANDARD; PRT; 1225 AA.
 AC Q90640; Q90608;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chromosome-associated kinesin KIF4A (Chromokinesin).
 GN KIF4A.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]

SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
 RP STRAIN=White leghorn; TISSUE=Embryonic retina;
 RX MEDLINE=95181533; PubMed=7876303;
 RA Wang S.Z., Adler R.;
 RT "Chromokinesin: a DNA-binding, kinesin-like nuclear protein.";
 RL J. Cell Biol. 128:761-768 (1995).
 RN [2]
 RP SEQUENCE OF 728-1088 FROM N.A.
 RC STRAIN=White leghorn; TISSUE=Embryonic retina;
 RX MEDLINE=94151328; PubMed=8108415;
 RA Wang S.Z., Adler R.;
 RT "A developmentally regulated basic-leucine zipper-like gene and its expression in embryonic retina and lens.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:1351-1355 (1994).
 CC -1- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR SPINDLE STABILIZATION.
 CC -1- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic chromosomes.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN PROLIFERATING CELLS;
 CC -1- NEUROEPITHELIUM OF EMBRYOS.
 CC -1- SIMILARITY: Belongs to the kinesin-like protein family. Chromokinesin subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U18309; AAC59666.1; -;
 DR EMBL; U04821; AAA18960.1; -;
 PIR; A56514; A56514.
 DR HSSP; P17119; 3KAR.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; DNA-binding;
 KW Nuclear protein; Coiled coil.
 FT DOMAIN 1 351 KINESIN-MOTOR.
 FT DOMAIN 352 1003 COILED COIL (BY SIMILARITY).
 FT DOMAIN 1004 1225 GLOBULAR.
 FT NP BIND 88 95 ATP (POTENTIAL).
 FT CONFLICT 1087 1088 KG -> RI (IN REF. 2).
 FT SEQUENCE 1225 AA; 138923 MW; FA01ED83425F5875 CRC64;
 SQ
 Query Match 6.9%; Score 917.5; DB 1; Length 1225;
 Best Local Similarity 25.2%; Pred. No. 6.2e-18;
 Matches 344; Conservative 242; Mismatches 428; Indels 349; Gaps 54;
 QY 3 EGVAVCVVRVPLNSREESLG-ETAAQVYVKKTDNNVYQVDSKSFNDFRPHGNETKN 61
 Db 6 EKGIVRVVRCPPLVPKTESGCQWCLSFVGPQVI--VGSDKAFYDYVDFSPSEQEE 63
 QY 62 VYEEIAAPLIDSAIQYNGTIFAYQTASGKTYTMWG-----SEDH---LGVIPRAHDLIF 114
 Db 64 VENTAVAPLIRIGFKGNATVLAYQTGSGKTYSMGTYTASQEHDPMSMGVPRVTKLDF 123
 QY 115 QKIKKPPDFRFLRVSYMEIYNTITDLCGT-QKMKPLIREDVNRNVVADLTVEVY 173
 Db 124 KEKEQFQDWFEVLKVSYLEIYNEIDILLCSSRERSQSISIRDPKEGKIVGLTERNVA 183
 QY 174 TSEMALKWLTKEKSRHYGETKMNORSRSRSHIFRMILSRKESGEPSCGSKVYKSHNL 233
 Db 184 SARDTVSCLEQGNCRVTASTANSSSSSHALFTICIDQKKK---NDKNSFSFKLHL 239
 QY 234 VDLAGSERAQOTGAAGVRLKEGCNINRSLFILGQVTKKL-SDQVGVGFNYRSLKTRIL 292

Db 240 VDLASERQKTKAEGDRLEKGINNGLICLGNVISALGEENKKGFPYRDSKLTLL 299
 QY 293 QNSLGGNPKTRIICTTPV--SFDETILALQFASTAKYMKNTPVNEVSTDEALLKRYK 350
 Db 300 QDSLGGNSHTLMTACVSPADSNLEETLNTLYRADRKIKNKPIVN-VDPQAEHLNHLKQ 358
 QY 351 EIMDLKQLEE-----VSLTRAQ-----MEKDLAOLLEEKDLLOK-----VQNEK 393
 Db 359 QVOOLQVLLQAHGGTLPVINSMAFSENLSLMEKNQ--SLMEENEKLSRGLSEAGQT 416
 QY 394 IENLREMLVTSLSLTQLQELKAKRKRVWCLGKINKMKNSYADQENIPTNTTTHKL 453
 Db 417 AQMLELIVTEQE--NEKNAKLEQQAQHAVCKLIDLOK----- 452
 QY 454 SINLLREI-DESVCSDDVFSNTDLTSEIE-----WNPATKLLNOENIESELNSLRAD 506
 Db 453 ---LLETVEDEELKENVEVRNIQQVLAQFQSESAAAEATEMANAE----- 497
 QY 507 YDNVLVDYQLRTEKEEMELKLEKNDLDEFALERKTKKQDQEMQLIHRISNKNIVKHR 566
 Db 498 -----QDAAGEAETGVTRKSSDDFTTQHAL-RQAQMSKELVELNKALALKEALAK 548
 QY 567 EYVNOQLENSKVELLEKEDQIKLOFYIDSOQLENIKMDLSYLSIESIEDPKMKQT 626
 Db 549 MIQN-----DSQLEPIQ-----SOYQTNIK-DLELEVSNLQ----- 578
 QY 627 LPDAETVALDAKRESAFLSENLEKEMKELATTYKOMENDIQLYQSOLEAKKMQVDL 686
 Db 579 -----KEK-BELILALSMKQDVN--QAKLSERR----- 605
 QY 687 EKELQAFNEITKLSLIDCKVPKLLCNLEEGKITDLOKELNKEVEENALREEVILL 746
 Db 606 -KRLQ-----ELEGQINELKXKLN-----EQAKLL 629
 QY 747 SELKSLPSEVERLRKEIODSEELHIITSEKDKLPSEVVHKSRSVQGLLEIGTKYDULA 806
 Db 630 KLESTERTVSKLQEIHEM-----KNQVQ-LMRQ----- 660
 QY 807 TTQSNYKSTDOFQNFKTHMDFEQRYKRV--LEENERNQ--EIVNLSKEAQKQPSLSICA 863
 Db 661 -----KEDAEKFRQWK-----QQKDEVIQKEDRKQVYELLKLEDFQK-QASVLR 707
 QY 864 LKTELS-----YKQLEKTRERVRQRLNE--MEQLKEOLENRDPSLOTVEREK 910
 Db 708 RKTETAAANKRLDALOKQREAADRKESQNRGMEGVAARVKSMLANEVEVLVSTERAR 767
 QY 911 TLITEKQQLLEEVKTLTOEKDOLKQLESQIERDQKLSHDHVTNNNIDTQEQRLNAL 970
 Db 768 ---RHLADLLEDKRLAQELLQLEKESGENPPSKLRRRTYSITDIQASEMDL----- 818
 QY 971 ESKQHQBTINTLKSITSEVSRNLHMEENTGETKDFQKQKMGVIGDKKQDLEAKNTQTLT 1030
 Db 819 -SLSKQIESLET-----EMELRSAQIAD-LQKLLDAD----- 849
 QY 1031 ADVKDNEIIEOQRKIFSLIQKNEHLQOMLESVIAEKEQKTDLKENIEMTIENQBEFLRL 1090
 Db 850 ---NGDRVKQRWDNIATILEAKCALKYLLGELVSSKVQ-ESKLESSLQKSTNCSDI--- 902
 QY 1091 GDELKKQOEIVAOBKNAHKEGELSRTCDBLAEEVEKLEK-----SQOLEKQOQLLNV 1146
 Db 903 -----QKMLIERNHATEAEAFQ---NOLLQEQHQOQEVLYLLSQFOQKAPAGKV 952
 QY 1147 QEEMSEMQKINEIENLKNLKNKELTLHEMETERLELAQKLNENYEVKSIKERVYK 1206
 Db 953 EDSLSEQEKQOE-----RLKFEKELEKVR-----EICEKNQELLOE-NDVLKQKMLLV 1001
 QY 1207 ELQKSFETERHLRGYREIATGLQVKEELKIAHIHL-----KEHQETIDELRRSVSEK 1261
 Db 1002 QVASGQKLRDQ-----QVSPSPDQPFQVIPPKTRRQTVAKPRAPTE- 1047
 QY 1262 TAQINTQDLKSHTKLQETIPVHLHEQELLPNVKKYSETQET 1304
 Db 1048 ---MNVEELFSDSESGEE-----EBAEWVP-VKAAGTKKS 1080

RESULT 14

SPOF SCHPO

ID SPOF SCHPO STANDARD; PRT; 1957 AA.
 AC Q10411; Q9USE9;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Sporulation-specific protein 15.
 GN SPO15 OR SPAC1F3.06C.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
 RX MEDLINE=20107136; PubMed=10639340;
 RA Ikemoto S., Nakamura T., Kubo M., Shimoda C.;
 RT "S. pombe sporulation-specific coiled-coil protein Spo15p is localized
 RL to the spindle pole body and essential for its modification.";
 J. Cell Sci. 113:545-554(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Young R.G., Aert R., Robben J., Grimmonprez B.,
 RA Wellens J., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer C., Holzer H., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 Nature 415:871-880(2002).
 RN [3]
 RP SEQUENCE OF 705-871 FROM N.A.
 RC STRAIN=968 h90;
 RX MEDLINE=20223868; PubMed=10759899;
 RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
 RA Hiraoka Y.;
 RT "Large-scale screening of intracellular protein localization in living
 RL fission yeast calls by the use of a GFP-fusion genomic DNA library.";
 Genes Cells 5:169-190(2000).
 CC -!- FUNCTION: Has a role in the initiation of spore membrane
 CC formation.
 CC -!- SUBUNIT: Monomer.
 CC -!- SUBCELLULAR LOCATION: Spindle pole body.
 CC -!- SIMILARITY: Belongs to the MPC70 family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; Z70690; CAA94624.1; --
 DR EMBL; AB027811; BAA87115.1; --
 DR PIR; T38077; T38077.
 DR GenDB SPombe; SPACIF3.06c; --
 KW Sporulation; Coiled coil.
 FT DOMAIN 199 785 COILED COIL (POTENTIAL).
 FT DOMAIN 804 1235 COILED COIL (POTENTIAL).
 FT DOMAIN 1320 1471 COILED COIL (POTENTIAL).
 FT DOMAIN 1481 1723 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

Query Match 6.7%; Score 893.5; DB 1; Length 1957;
 Best Local Similarity 20.3%; Pred. No. 4e-17;
 Matches 425; Conservative 448; Mismatches 800; Indels 419; Gaps 76;

212 ESREKGEPSNCEGSKVSHNLNVLADGSERAAQTGAAGVRLKEGNCNINRSLFILGQVVK 271
 62 DSSEELKROEVRGRRHSDLSIDAKLGSSEGS-TASSALPLTPRSPNASWLLVR----- 115
 272 LSDQGVGFNYRDSKTRILONSLGNPKTRICTTPVSFDETLTALQFASAKYMKN 331
 116 -----GGLDPSILDINSVTQKSNLLNELKQVASKLAALHEHNGIISLQSSNKKDKN 169
 332 TPVNEVSTDEALLKRYK-----EIMDLKKOLEEVS--LETRAQAMEKDQ- 375
 170 TSSVTITSEEDVSFYOKLQNMESNPSAKOSEAYDLRSQLLITVEKLDKKEKYDIKE 229
 376 -----LAQLLEKDLQKQVNEKIENLRLMLVTSSEL-TLQQ-----ELKAKRRVTVCL 425
 230 DVSIKASLAEOASNSLRGEQ-ERLEKLVSSNKTVSTLRQENSIRABCK----- 281
 426 GKINKMKNVADQFNPTNITTKHLSIN-----LLREIDESVCSSESVFSTLPTLS 480
 282 -----TLQEKLEKCAINEEDSKLEELKHNANVANSDAIVHK-DKLI 321
 481 ELENPATKANQENIELMSLRADYDNL-----VLDVEQLRTEKEEMELKKE 530
 322 E-----DLSTRISFEFNLSKERTLSTKNEKLEKLRLNTIGSLKDSKTSNLSOLEEWEVLKE 378
 531 KNDLDEFBALRKRKTKDOEMQLIHEISNLKLVHREYVNOENELSSKVELLREKEDQ 590
 379 SNRIHSQLTDAESKLSFEQ---ENKSLKSGI-----DEYQNNLSKDKMVKQVSSQ 428
 591 IKKQVEYI--DSQKLENIMDLSYSLESIEDPKQMKQLFPAETVALDAKESAFRSEN 648
 429 LEEARSLAHATGKLAENISERDFQNKIKQFEKTEQDL-----RACLNSSS 475
 649 LELEKMKELATTYQMENDIQLYQSLEAKKMQVDLEKELQSAFNEITKLITSLIDGKV 708
 476 NELKEK-----SALIDKQOELANLRQIKQKVSSTQSSLOSORDI-----LNEKK 525
 709 PKDLNLELEKGTIDLOKELNKEVEENFALREEVILLSELKSLPSEVERLKEITQDKSE 768
 526 KHEV-----YESQLNELKGEQTETNSSEHLS-----SOLSTLAAEKEAAVATNNELSE 574
 769 ELHIITSEKDKLPFSEVHVHESRVQGLLEIGTKTDDLATQSNYKSTDOEFQNFKTLMD 828
 575 -----SKNSI-----QTLNAPQEKIAXSMQMKENQNFSSLDTSFKKUNESHQ 620
 829 FEQYKVMLEENRMNOEIVNLSKEAQFDSISGALKTELSYKTOBLOKTRREVORLNE 888
 621 LENNHQTTIKQLKDTSSKLOQLERANFEQ-----ESTLSDENNDRTKLLKLBESNKS 676
 889 MEQLKEQLENRDSPLQVREKTLITEKLOQLLEEVKTLTQEKDDLKQLESQIQRDOL 948
 677 LIKKQEDVDSLEKNIQTLKEDLRKSEALRFSKLEAKNLRVIDNLRKGHETLEAQNRDL 736
 949 KSDIHTVANNIDTQQLNLALESKQHOETINTLASKISSEVSRNLHMEENTGETKDEF 1008
 737 HSSLSDAKNTNAILSLSELTKSSSDVXRLTANVETL--TQDSKAMKQSFSTLVSYSQISNL 795

1009 QOKMVGIDKKQDLEAKNTOTLTADVK--DNEIIEOQRKIFSLIOEKNELOOMLESVIAE 1065
 796 YHEL--RDDHVNMQSQNNITLSESKLTKTDCENLTQON--WILI---DNVQKLMKHVQ 848
 1066 KEQKTTDLKE--NIEMTIENOEBELRLGDELAKQOQEIIVAQEKXHAIKKEGELSRCDRL 1122
 849 ESKV-SELKEVNGKLSLDLKN-----LASSLNVAISDNDQILTQALAEKSNYDLS 897
 1123 AEVEEKLKESKQOQKQOOLNVQEMSEMOKKIN--EENLKNELKNKELTLEHMET 1179
 898 EQESAQLNSGLKSLAEKQLLHTENBELHRLDKLTKLIEESKSDGLKGLKTARQOEI 957
 1180 ERLELAQKLENVEEVKSIKERKVLKE-LQKSFETERD--HLRGYRREIETAGT---LOT 1233
 958 SNLK-----EENMSQQAITSVKSKLDETLSKSKLEADIEHLKNKVSVEVERNALLAS 1012
 1234 KEBLKIAHILKHEQHOETIDELRRSVSEKTAQIINTQPLEKSHTKLQOEIPVLHHEQE--L 1291
 1013 NERLM--DDLKNGENIASLQTEIEKRAE-----NDDLOSKLSVSVSEYENLL 1059
 1292 LPNVKVSQETQWNELELLTBOSTTK--DSTTLARIEMERL-----RLNEKQESOEI 1344
 1060 LISSQTNKSLDKTNQLKXI--EKNVQKLDEKQORNVLEBELTSKYKGLKEENAIQIDEL 1118
 1345 KSLTKER-----DNLKTIKEALEVKHQKEHIRETLAKIQESQSKQOSLNMK 1393
 1119 LALRKSKKQHDLCANFVDDLKESDALE---QLTNEKNELIVLSQSNNEA---LV 1171
 1394 EKONETTKIVSEMEQ-FPKPKDSALLRIEIMGLSKELQESHDMKSVAKEDDLO-RLQ 1451
 1172 EERSDLANRLSDMKKSLSDSDNVISRLVRV-----NDELDTLKDKKDSLSQTS 1224
 1452 EVLQESDQKLENKTEIVAKHLETEBELKVAHCCLEKEQFETINELRVNLSKEKTEISTIQ 1511
 1225 EVCQDRDLDLS-----LKGESEFNKYAVSLRELCTK----- 1257
 1512 KQLEAINDKLQNKQTEIYEEKEQNLKIQISEVQE--NVNLEKQFKEHRKAKDSALQSIRESK 1570
 1258 -----SEIDVPVSEIIDDNFVFNAGNFSLSRLTVLSLENYLD-----AFNQVNFK 1303
 1571 MLELTNRLOESQEBIQIIMKEEEMKRVQEAQIARDQKLENTKEIIVAKMESQEKQFQ 1630
 1304 KQELDNRLTTTDAFTKVVADLEKLOQEHDDWLIQDGLD-----KALKDSERNF 1353
 1631 LKMTAVNTEQEMCEIEHLKEQFQTKLENIENTENIRLTQILHENLEEMRSVTKRDD 1690
 1354 LRKEA--EMTENIHSLEEGKEETKEIAELSSRLDNQIATNKLKNQDLHL-----NOE 1405
 1691 LRSVEETLKVREDQKLENIRLETITRDLEKQBELKIVHMLKHEQETIDKLRGIVSEKTN 1750
 1406 IRLKEDVLK-EKESLIISLESLSNQROKESL-----LDKNE 1443
 1751 ISNMOKDLEHNSDALKAQDLKIQEBELRIAHMLKEQQETIDKLRGIVSEKTKDNQKMD 1810
 1444 LEHMLDDTSRKNSSLMEXIESINSSLDKSPELASAVEKLGALQKLSHESISLMEINISQ 1503
 1811 LENSNAKIQ--EKIQELKANEHQILTLKQVNETQKVVSEMEQLKQKQKQDQSLTLKLE 1867
 1504 LQEAKEKIQVDESTIQEL---DHEITASK--NNTYKGLD-----KD----- 1540
 1868 TENINLAQELHENLEEMKSVM-KERDNLRRVEETLKLERDQKLESIQETKARDLEIQOEL 1926
 1541 ----SIIIDLSINIQOLNNLLAEESKAVKLSSEKSELSLQFNLSRLADLEYHKSQVSESL 1596
 1927 KTWARM--LSKEHKEVTDKLREKISIKTIQISDIQKDLKSDKDELOKKIQELQKQELQL 1983
 1597 GRSKKLASTTEELQLAENERLSLITRMLDQNVKDLQKLSNIDKLSIEDLRTLSRSDSVA 1656
 1984 RVKEDVNMHKKINEMEQKQFEPNLYCKCEMDNPFQTLKKLHESLEETIRIVAKERDE-- 2041
 1657 SLQKECKI---KSNTVESLQ-----DVLTSVQARNAEDEVSRSVDKIR---RRDRC 1704

QY 2042 -----LRRIKESIKMERDQPI-ATLREMIARDQNHQVKEPKRLSDGQOHLMSLREK 2094
 Db 1705 EHLGKLLKJHSLQEQHETFFRAEQQMTQGLFKETVKKQEK-----LKKL 1753
 QY 2095 CSRIKELLKRYSEMDHYECINRLSLDLKE-----IEPHRMKKLVLSVYTK 2144
 Db 1754 NLQEQILIPASSIL--VYESYIR---DIEKEIIVLOERLNGIELSQQLPKGYGYPFKT- 1807
 QY 2145 IKEQHECINRPMDFIDEVEKO-----KELLIKIHOLOQDCDVPSE 2187
 Db 1808 -----NRVMEVLDSPKQVAKIQFLAGAEFIVKFEDLEKCAAEKE 1850

RESULT 15
 MFLP1_YEAST
 ID MFLP1_YEAST STANDARD; PRT; 1875 AA.
 AC Q02455;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin-like protein MFLP1.
 GN MFLP1 OR YKR095W OR YKR415.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c;
 RX MEDLINE=93247549; PubMed=8483450;
 RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
 RT "A new yeast gene with a myosin-like heptad repeat structure.";
 RL Mol. Gen. Genet. 237:359-369 (1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94205265; PubMed=8154196;
 RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
 RA Renacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
 RT "The complete sequence of a 15,820 bp segment of Saccharomyces
 cerevisiae chromosome XI contains the UBI2 and MFLP1 genes and three
 new open reading frames.";
 RL Yeast 9:1349-1354 (1993).
 CC -!- FUNCTION: Myosin-like protein that is probably involved in DNA
 CC repair.
 CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
 CC -!- CAUTION: Ref.2 misquotes the gene name as "MPL1".
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; L01992; AAA34783.1; -;
 DR EMBL; X73541; CAA51948.1; -;
 DR EMBL; Z28320; CAA82174.1; -;
 DR PIR; S38173; S38173.
 DR Germline; 140074; -;
 DR SGD; S0001803; MFLP1.
 DR GO; GO:0005635; C:nuclear membrane; IDA.
 DR GO; GO:0005654; C:nucleoplasm; IDA.
 DR GO; GO:0006606; P:protein-nucleus import; IDA.
 DR KW Coiled coil; DNA repair.
 FT DOMAIN 69 487 COILED COIL (POTENTIAL).
 FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
 FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
 FT CONFLICT 301 301 R -> A (IN REF. 1).
 SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C9066867 CRC64;

Query Match 6.7%; Score 891.5; DB 1; Length 1875;
 Best Local Similarity 20.7%; Pred. No. 4.4e-17;

Matches 469; Conservative 467; Mismatches 756; Indels 579; Gaps 96;
 QY 462 DEVCSESDVFSNTLDTLSIEWNPATKLINQENIESELSNLRADYDNLDVDEQLRTE- 520
 Db 18 DERLNAIAFFGCSLQVKSFD-GDVVHLLNDKLL--QFNELKSENKLVTVSDFELKASS 74
 QY 521 -----KEEMELKLEKNDLDEFALERTYKQDQEMQLIHEISNLKLVKHREVNQDL 573
 Db 75 LKIDGLKTEMENVIRENDKI-----RKERND-----TFVKFESVENEKM 114
 QY 574 ENELSSKVELLREKEDQIKLOEYIDSOXLENLKMDSLSYLSIE-----DPKQ 622
 Db 115 --KLSELEFVVRKLDLDEEKKETOSNOORTUKI-LDERLKEILVRVNNNSNBECK 171
 QY 623 MKQTLFDAET-----VALDAK-RESAFLRSNLELKEKMK--ELATTYKQ- 664
 Db 172 LRSTMIDLETQGYITNDLNGSRTELETKTQELTLQSNNDWLEKLSKNEQYLSYRQK 231
 QY 665 -----MENDIQLYQSLEAKKQVQVLEKLSQAFNITKITLSIDGKVPK 710
 Db 232 TDKVILDRNELNRLNDFQMERITNDVLKKNNELSKLSQELKLEIKGLSDSLNSE-KQ 290
 QY 711 DLLCNLELGKITD-LOKELNKEVEENEALRE---EVLILSELKSLPSEVERLRKEIQK 766
 Db 291 EFSAEWSLKQRLVDLLESQUNAVKELSIRELNTAKVIADDSKKQTPENEDLLKELQLT 350
 QY 767 SEELHIITSEKDKLFSVVRHVESVQGLLEETGKTKDDIATQSN-----YKST 815
 Db 351 KEKL--AQCEKEL-----RLSSITDEADENENLSAKSSDDFLFKQLIKERT 399
 QY 816 QDEFON-FKTLHMDPEQYKVVLENERMNOEIVNLSKBAQKFDSSLGALKTELSTYQOE 874
 Db 400 KEHLQQLQETTFIVELEHKVPIINSFKRTDM-----LENELNAAILL 441
 QY 875 LQEKTRVQOERLNEMLQLEOLENEDSDPLOTVEREKLITLTKLOTLBEVXTLTQEKDDL 934
 Db 442 LEHTSNEKNKAVKELNAXKQKLVCENDLQTLTKQRLDLQCIQYLLITNSVNSDKGL 501
 QY 935 KQLQESLQIERDQKSDIHTVNMNIDTQEURNALESKQHOETINTLKSISB--SVS 992
 Db 502 R--XEEIQFQINQED--DSTITESDSQKVVTRELVEFK---NIILOEQKNAELLKV 553
 QY 993 RNLHEENTGETKDFEQKQVGDKKQDLEAKNTQTLTADVKNDEIIEQQKIPFSLIOEK 1052
 Db 554 RNL-----ADKLESKEKSKQSL-QKIESETVNEAKEALITLKEK 593
 QY 1053 NELOQMLESVIAEKEQLKTDLKENIEMTIENQEBELRLGDELKKQOEIVAEKNAIKKE 1112
 Db 594 MDLESRIEELQKELEELKT-----SVNED-----ASYSNVTIK-- 627
 QY 1113 GELSRTCDRLAEVEBEKLEKSKSQQLEKQOQLNVQ--EEMSEMOKKINIEKLNELK-- 1168
 Db 628 -----QLTETKRDLESQVQDLQTRISQITRSTENMNSLLNKEIQDLYDSKSDISKL 679
 QY 1169 NKELTLEHMETERLE-LAQKLNENYEVKSITK-----ERKVLKELQKSFTEHDHL--R 1220
 Db 680 GKEXSRILAEERFKLLSNTLDTLTKAENDQLRKRFQYDLQNTILKQDSKTHETLNEYVSK 739
 QY 1221 GYTRIEIATGLQTEELKIAHILKHOETIDELRRSVSEKTA--QIINTODLEKSHYK 1277
 Db 740 SKLSIVETELLNLKEBQKL-RVHLERKLNK--QELNKLSPKDSLRINMVTLQTLQKRED 796
 QY 1278 LOBEIPVLHBEQELLPNVKVSETQETMNELELLTQSTTKDSTTLARIEMERLRINEKF 1337
 Db 797 LLEBTRKSCQ-----KKIDELEALSSEL-----KKETSQKDHIIKQLEDN---NSNI 841
 QY 1338 QESQEEIKSLTKERDNLKITKEALEVHKHOLKEHIRETLAKIQESQSKQPSQSLNMEKON 1397
 Db 842 EWTQNKIEALKDYEVSVITSVDKQTDIEKLYQKVKASLEKEIEEDKIRLHTYNNVMDETIN 901
 QY 1398 ETTTKIVSEMQFKPKDSALLRIEIMGLSKRLQESHDEMKSVAKEDDQLQRIQEVLOSE 1457
 Db 902 D-----DS---LRKELEKSKIN--LTDAYSQIK---EYKDYLETTSQSLLQQT 940

QY 1458 SDQLENKEIYAKHLETEBELKVAHCCLKEQETINELRNLSEKET-----EISTIOKQ 1513
Db 941 NGLKDESPKDF-----NQIK-NUTDEKTSLEDKISLLKEQ 975
QY 1514 LBAINDKLQNKQIEIYEKEEQLNIQISEVQENNELKQFKEHRKAKDSALQSIESKMLE 1573
Db 976 MFNLNNELDLQ-KKGNEKEKADPKRISILQNNNKEVEAVKSEYKSLKIQINDLDQOTI 1034
QY 1574 LTNRLQESQEEIQTIMKEKEMKRVOR-----ALQIERDOLKENTKEIVAKM 1620
Db 1035 YANTAQNVE--QELQKHADVSTISLREQLHTYKGVKTLNLSRDQLENALKE----- 1087
QY 1621 KESQEKYQFLKMTAVNETQKMCETHELKEQFQTKLNLNETENIRLTQILHENLEE 1680
Db 1088 ---NEKSWSSQK-----ESLLEQLDLSNRIEDLSSQN---KLLYDQIQI 1126
QY 1681 MRSVTKERDD-----LRSVEETLKVVERDQLENLRETIT-RDLEK-QBELKIVHMLKE 1732
Db 1127 YTAADKEVNNSTNGPGLNNILITLRRERDIL--DTKVVAERDAKMLRQKISLMDVELOD 1184
QY 1733 HQBTIDKLRGIVSEKNEINMQKD---LEHSN--DALKAQDLKIQELRAIAHMLKEQ 1787
Db 1185 ARTKLDNSR---VEKENHSSIIQOHDIMEKLNQNLRESNITLRNELENNNNKKKBLQ 1241
QY 1788 ETIDKLRGIVSEKTDKLSNMOKDLENSNAXLOEKIOELKANEHQIITLKDVNETQKVS 1847
Db 1242 SELDKLK-----QNVAPIESELTAALKYSQWKEQELK-----LAKEEVHWWKRSQ 1287
QY 1848 EMEQLKQIKDQSLTSLKRIENLNLQAEHLNLEEMKSVNMKRDN-LRRVEE-----TLK 1902
Db 1288 DI--LEKHEQLSSDYKLESEIENLKEEL-ENKEROGAEEAEKFNRLRQAOERLKTSK 1344
QY 1903 LERDQLKE-----SLQETKARDLEIQ-----QELKTARMLSKHEKETV 1940
Db 1345 LSQDSLTEQVNSLRDAKNVLENSLSEANARIEELQNAKVAQGNQLEAIRKLQEDAERAS 1404
QY 1941 DKLEKISEKTIQ---ISDFOKDLKSKDLQKKIOELQKELQALLRVKEDVNMHHKI 1996
Db 1405 RELQAKLEESTSTVESTINGLNEBHITTLKEIEKQRIQQ--QLQATSANEQNDLS--- 1458
QY 1997 NEMQLAKQEPFNLYCKCEMDNFQITKKLHESLEIRIVAKERDELRIKESLKMWDQF 2056
Db 1459 NIVESMKKSF-----EEDKIKFIKKTQ-EVNEK 1486
QY 2057 IATLREMTARQNHQVPEKRLISDGOHLMESLREKCSRIKELLKEYSEMDDHYECLN 2116
Db 1487 ILEAQERLNQPSNINMEEIKKKWESEHQVVSQKIRE---ABEALKK----- 1530
QY 2117 RLSLDLEKEIEFHRIMKLVLSVVTIKIKEQHECINKFEMDFIDEVEKOKELLIKIQH 2176
Db 1531 KIRLPTEEKI-----NKIIEKKK--ELEKEPEEKVEE-----RIKS 1565
QY 2177 LQQCDVPSRELRLKLNQNDLHTEILDKSESEPPSIKTEFOQVLNKRKEMTQFLEE 2236
Db 1566 MEQSGEI-----DVVLAKQLEAKVQEKQKEL-ENEY-----NKK-----LOE 1601
QY 2237 WLNTRFDLEKLNKGTCQENDRICQVNNFNFNRIIIMNTESTEFERSATISKWEQDLKS 2296
Db 1602 ELK--DVPHSSHISDDERDKL-----RAIESR-----LREFPNNELQA 1638
QY 2297 LKEKNEKLFKNYQTLKTSLASGAQVNPFTQDNKNPHVTSRAQTLTTEKIRELENSLHEAK 2356
Db 1639 IKKSFDEGKQAMMKTTILE-----RKLAKMESQLSETK 1673
QY 2357 ESAMHKESKIIIRKQKELEVTNDIIIAKQAKVHESNKC-----LEKTYETIQVLQDKVALGA 2412
Db 1674 QSAESPSPKSVNNVQNP-----LGLPRKIEENSFPNPLISGEKLLKINSKSSGGF 1726
QY 2413 KPYKEIEDLKWKLKIDLEKMKNAKEFEKEI-----SATKATVEYQKEV 2457
Db 1727 NPFTSPSPNKLQNDKNDKRESLANKTDPPTHLEPSPNIPASRGLISSSSTLSLTDITNDEB- 1785

Search completed: July 29, 2004, 09:36:12
Job time : 52.5784 secs

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.
 OM protein - protein search, using sw model
 Run on: July 29, 2004, 09:39:47 ; Search time 133.008 Seconds
 (without alignments)
 6280.361 Million cell updates/sec
 Title: US-10-045-631B-88
 Perfect score: 13329
 Sequence: 1 MAREGAVVCVRPLNSRE.....SQPMPHASSGKDVPCKTQ 2663
 Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues
 Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000
 Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : Published Applications_AA.*
 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap.*
 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pap.*
 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pap.*
 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap.*
 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pap.*
 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap.*
 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap.*
 8: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap.*
 9: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap.*
 10: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pap.*
 11: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pap.*
 12: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pap.*
 13: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap.*
 14: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pap.*
 15: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pap.*
 16: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pap.*
 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pap.*
 18: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pap.*
 19: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				SUMMARIES	
Result No.	Score	Query Match	Length DB ID	Description	
1	1234	9.3	1382 16	US-10-437-963-176714	Sequence 176714,
2	1183	8.9	240 10	US-09-893-519A-67	Sequence 67, Appl
3	1125.5	8.4	3225 16	US-10-408-765A-254	Sequence 254, App
4	1077.5	8.1	1388 12	US-10-332-089-2	Sequence 2, Appli
5	1077.5	8.1	1388 14	US-10-146-473-82	Sequence 82, Appl
6	1077.5	8.1	1388 15	US-10-173-999-32	Sequence 32, Appl
7	1077.5	8.1	1388 16	US-10-188-832-164	Sequence 164, App
8	1064	8.0	3899 14	US-10-171-311-4	Sequence 4, Appli
9	1064	8.0	3917 14	US-10-171-311-8	Sequence 8, Appli
10	1058	7.9	3907 14	US-10-171-311-2	Sequence 2, Appli
11	1058	7.9	3925 14	US-10-171-311-6	Sequence 6, Appli
12	1055	7.9	3911 15	US-10-370-685-100	Sequence 100, App
13	1055	7.9	3911 16	US-10-408-765A-1839	Sequence 1839, App
14	1052.5	7.9	3878 14	US-10-080-608A-11	Sequence 11, Appl
15	1041.5	7.8	2383 14	US-10-082-830-260	Sequence 260, App

16	956.5	7.2	1232	14	US-10-116-712-670	Sequence 670, App
17	956	7.2	1232	16	US-10-408-765A-2153	Sequence 2153, Ap
18	956	7.2	1235	15	US-10-334-143-8	Sequence 8, Appli
19	944.5	7.1	1232	14	US-10-116-712-664	Sequence 664, App
20	944.5	7.1	1232	15	US-10-116-712-669	Sequence 669, App
21	944	7.1	1237	14	US-10-334-143-33	Sequence 33, Appl
22	938.5	7.0	694	12	US-10-425-114-59725	Sequence 59725, A
23	910.5	6.8	1401	16	US-10-287-226-142	Sequence 142, App
24	899.5	6.7	1392	16	US-10-287-226-86	Sequence 86, Appl
25	898.5	6.7	1392	16	US-10-473-574-27	Sequence 27, Appl
26	893.5	6.7	1392	16	US-10-369-493-2070	Sequence 2070, Ap
27	891.5	6.7	1875	15	US-10-369-493-22285	Sequence 22285, A
28	886.5	6.7	1905	15	US-10-369-493-5368	Sequence 5368, Ap
29	873.5	6.6	1827	15	US-10-661-809-23	Sequence 23, Appl
30	870	6.5	10203	16	US-10-370-685-105	Sequence 105, App
31	867.5	6.5	963	14	US-10-080-608A-20	Sequence 20, Appl
32	867.5	6.5	963	15	US-10-370-685-109	Sequence 109, App
33	867.5	6.5	2816	15	US-10-240-145-145	Sequence 145, App
34	866.5	6.5	957	14	US-10-080-608A-16	Sequence 16, Appl
35	866.5	6.5	957	14	US-10-370-685-105	Sequence 105, App
36	863	6.5	956	14	US-10-080-608A-17	Sequence 17, Appl
37	863	6.5	956	15	US-10-370-685-106	Sequence 106, App
38	856	6.4	967	14	US-10-080-608A-21	Sequence 21, Appl
39	856	6.4	967	15	US-10-370-685-110	Sequence 110, App
40	854.5	6.4	1979	14	US-10-205-823-419	Sequence 419, App
41	854	6.4	1805	10	US-09-820-843A-73	Sequence 73, Appl
42	854	6.4	1805	12	US-10-282-122A-63513	Sequence 63513, A
43	850.5	6.4	1394	16	US-10-287-226-84	Sequence 84, Appl
44	848.5	6.4	963	14	US-10-080-608A-22	Sequence 22, Appl
45	848.5	6.4	963	15	US-10-370-685-111	Sequence 111, App

ALIGNMENTS

RESULT 1

US-10-437-963-176714
 ; Sequence 176714, Application US/10437963
 ; Publication NO. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrei A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 176714
 ; LENGTH: 1382
 ; TYPE: EXT
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_74437C.1.pap
 ; US-10-437-963-176714

Query Match 9.3%; Score 1234; DB 16; Length 1382;
 Best Local Similarity 28.6%; Pred. No. 1.5e-53;
 Matches 432; Conservative 254; Mismatches 514; Indels 310; Gaps 54;
 QY 7 VAVCVVRPLNSREESIGETAGQVYWKTDNNVI-YQVDGSKSFNDRFVHGNETTKNYEE 65
 Db 4 IHVAVRPLTA--EDAGSP---WRVSGNAIALSTQPSIRFEDRIFGECRTADYGA 58
 QY 66 IAAPIIDSAQVNGTIFAYGOTASGTYTMCSEDLHGVIPRAIHDFOKIKKFPDP 125
 Db 59 RTKHIIVDSVRGNGTVFAYGQTNKGTMYRSGNEPGIPLAVHDLFTTIEHLDREF 118

1094 LKQBEIQAQENKHAKEGELSRCTDLRAEVEEKLKESQOLQEQOQLL-----1144
 1098 RSRFGCVGVGRGRTPRGGVDGWVACVAPNSSAAQWSSAGGEVAGVLRACAGGIWWH 1157
 1145 ---NQVEEMSEMKKINEIENKLNELNKLLELM-----E 1178
 1158 LVASVADTMFGWSEVSKALLGLPLMLATATPSGIVHLVEGVAIGALVOLHIKGLRTPFE 1217
 1179 TERLELAQK-----LNEN--YEEVKSIKER-KVKLELOKSFETERDHLRGYIREIATG 1230
 1218 SERLLAKERNIDTGVNENELHQOLLSITEERDKLSEIK-----YMSVINESE-- 1266
 1231 LQTKELKIAHLKHEQHTIDELRRSVSEKTAQIINTQDL-----EKS 1274
 1267 -----LIQAKATIDELSSRISIVEAKMKNYWLPLSNLGDCEFDASAYNKE 1313
 1275 HTKLOEIPVIEHEQELLPNVKVSQETQETMNELELLETE---OSTTKDSTTLARIEMERL 1331
 1314 NTKLRWQIRWQPELD-----AHRGLRKEAINEKMLMDTKYLEASTKLLKOLSPYCREVL 1368
 1332 RLNEKFOESQ 1341
 1369 RLKEQLKESQ 1378

RESULT 2

US-09-893-519A-67
 ; Sequence 67, Application US/09893519A
 ; Publication No. US20030027243A1

GENERAL INFORMATION:

APPLICANT: ANADYS PHARMACEUTICALS, INC.

APPLICANT: THOMPSON, Craig

APPLICANT: MOORE, Jeffrey

APPLICANT: BUURMAN, Ed T.

APPLICANT: BRADLEY, John

APPLICANT: DESILVA, Thamara

APPLICANT: HARRIS, Sandra

APPLICANT: KOMARNITSKY, Svetlana

APPLICANT: MENDILLO, Marc

APPLICANT: MOORE, Daniel

APPLICANT: MCCOY, Melissa

APPLICANT: SANDERSON, Karen

APPLICANT: HAQ, Tariq

APPLICANT: ZHU, Shuhao

APPLICANT: LONG, Fan

APPLICANT: DAVIDOV, Eugene

TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE

FILE REFERENCE: 0342/10548-US2

CURRENT APPLICATION NUMBER: US/09/893,519A

PRIOR FILING DATE: 2001-06-28

PRIOR APPLICATION NUMBER: US 60/215,164

PRIOR FILING DATE: 2000-06-29

PRIOR APPLICATION NUMBER: US 60/224,457

PRIOR FILING DATE: 2000-08-10

NUMBER OF SEQ ID NOS: 146

SOFTWARE: Patent in version 3.1

SEQ ID NO 67

LENGTH: 240

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

OTHER INFORMATION: Corresponds to SEQ ID NO: 140

PUBLICATION INFORMATION:

DATABASE ACCESSION NUMBER: Human Genbank/CAA78727

DATABASE ENTRY DATE: 1993-01-10

RELEVANT RESIDUES: (1) ..(240)

US-09-893-519A-67

Query Match 8.9% Score 1183; DB 10; Length 240;

Best local similarity 100.0%; Pred. No. 6.1e-52;

Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

126 LLRVSMYIYNETITDILCGTQMKPLIIPEDVNRNVYVADLTETEVVYTSMAKWMITKG 185
 119 LLRMSYIYNEEINDLL--VPEHRKQIHESIERGIYVAGLRBEIVTCEQVLEFMSFG 176
 186 EKSRHVGTEKMQRSRSHITFIMLESREKEGPPS-----NCEGSVKVSHNLVDLAGSER 241
 177 ESRHHIGETNMVYSSRSHITFRVIESREKVDSEAGSCD-AVRVSVNLVDLAGSER 235
 242 AAGTGAAGVRLKGCNINRSLFLIGQVKKLSDG--QVGGFINRYSKLTIRIIONSLGNN 239
 236 AAKTGAEGVRLKESHINKSLMTLGTIVIKLSGIEGQGGHPYRDSKLTIRIOPALGNN 295
 300 PKTRITICTIT--PVSFDEITLALQFASRALKRNVVNVSTDEALLKRYRKEIMDLKK 357
 296 ANTAIICNITLAQHADETKSSLOFASRALRVNCACVNBILDAALLKQREIEHLRA 355
 358 QL-----EVSLETRAQAMEKQDQAQLLEKDLQKVQNEKIEENLFRMLVTSLSLTLQ 411
 356 KLRSELEKERISLE-----LEBEKAKEQDKRLIE--QAKKIENL-----SSLVINS 401
 412 E-----LKAKRKRVRVTCIGKINMKNSVADOFNPTNITTKHKLINL---LREID 462
 402 ERDRTTVSSKNKRLTWCPLLSRQDQGVLESVQBDPPSSTVVRGRNMEMPLHFELI 461
 463 ESVCSSES-----DVFS-----NTLDTLSE-----481
 462 QSCSESSIKHYTDAYSGLSCEDDSLPDSHALLHVTSRRKPNMTKKSQDQQLMGLASER 521
 482 -----JEWNPATKLLNQENIESINSLRADYDNLVDYEQLRTEKEMEL-KUKERNDL 534
 522 IIPQELNDKWTYQ--SQENIKACVNGLSARESEAILVIKQLEDQIKLLEKESFPQNL 579
 535 DEFALERTKTKDQEMQIHIENLKNLVKREVNQDLENELSKVLELLREKEDQIKKL 594
 580 D--DVLELAT--QOKASFHE-----KYEELQONALVAQEQAKI--ANEKLSQEPAA 624
 595 QBYIDS--QKLENIKMDLSYSLESTEDPKQMKOTLFDATVALDAKRESAFILRSNLELK 652
 625 YEFLTGIFVETESIAVQDQSTRSDNALSFTEELF-----QNLFWAKNFTVS 673
 653 ERMKE-----LATTY-----KOMENDITQYQSLEAKKQVQVDLEKLOAFNHI 697
 674 IDNYOPFCSIYIDLVTGYGKPGKIGDRTITLAPAGLLLAELKEVEAK---QFVCGDI 730
 698 TKLTSLLD-----GKVPKDLLCNLELGHKIDLOKELNKEVEENALREEVILLSELKSL 752
 731 TQFSVIRDYENISNCLREKLSKLEMEKKILDEQSLDQKD-----ELQRL 775
 753 PSEVERLRKEIODKSEELHIITSEKDKLFSVVVHKESRV-----QGLLEEIGTKDDLAT 807
 776 KSSLESCEKAMEDCNQNEL-----EKDSIILSELLTLQEVYLSLSSSLMKESIRKELDR 832
 808 TOSNVKSTQDEFONFXTLHMFQEKYKWLVEENE-----RMNQEIWNLSKEAKQFDSL 861
 833 TTKLKETENKLN-----SIQEKILESEKAEQAQREIKLQSQRTLLERDLRKDSFT 886
 862 GALKTELSYKTOEL-----QKTRVQQLNEMEQIKQLENRDSPLQTVREKTLITEKLQ 918
 887 VDKRHEQSVKSELAGIYQAVQIOEDYGLKEMHAFDMEABIASLQ--EALVTIAEK-E 943
 919 QTLVEYKTLTOEKDQLKQLESIQIERDQLKSDIHDVTNNNIDTQQLRNAME-----SLK 974
 944 EALSRELLTSAVEDLESLNSAETSSILLEET-AVLTKLDAESISKKLEASISLS 1002
 975 QHOTTINTLKSISIEEYSRNLHMEENTGETYKDFQKQWGVGDKQDLKAKNTOTLADVK 1034
 1003 REKEDMGIELT-----DVLMESEGSTWTAK-----KAYLEAK--QKLNLCNK 1045
 1035 DN-EITTEQQRKIFSLIOEKNELQOMLESVIAEKEQLKTDLENIENTIEMIQEELRLIGDE 1093
 1046 NNCKLSBLLIKV-----QIQRISGGFRASPHWELRLRRRTAVVKAARGTVGSGPS 1097

QY 1801 TDKLSNMOKLNSNAKLOEKIOELKANEHQIITIKVDVNETQKVKVSEMEQKKQIKDQS 1860
 Db 1 TDKLSNMOKLNSNAKLOEKIOELKANEHQIITIKVDVNETQKVKVSEMEQKKQIKDQS 60
 QY 1861 LTLKLEINENLAQELHENLEEMKSVMKERDNLRRVEETKLKRDQIKESLOETKARDL 1920
 Db 61 LTLKLEINENLAQELHENLEEMKSVMKERDNLRRVEETKLKRDQIKESLOETKARDL 120
 QY 1921 EIQOELKTARMLSKHEKTVDKLREKISEKTIQISDIQKOLDKSDKDELQKKIQELQKKEL 1980
 Db 121 EIQOELKTARMLSKHEKTVDKLREKISEKTIQISDIQKOLDKSDKDELQKKIQELQKKEL 180
 QY 1981 QLLRVKEDVNVSHKKINEMEQKKQFEPNYLCKCEMNFQTLTKKLSLEIRIVAKERD 2040
 Db 181 QLLRVKEDVNVSHKKINEMEQKKQFEPNYLCKCEMNFQTLTKKLSLEIRIVAKERD 240

RESULT 3
 US-10-408-765A-254
 ; Sequence 254, Application US/10408765A
 ; Publication No. US20040101874A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ghosh, Soumitra S.
 ; APPLICANT: Fahy, Eoin D.
 ; APPLICANT: Zhang, Bing
 ; APPLICANT: Gibson, Bradford W.
 ; APPLICANT: Taylor, Steven W.
 ; APPLICANT: Glenn, Gary M.
 ; APPLICANT: Warnock, Dale E.
 ; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
 ; FILE REFERENCE: 660988.465
 ; CURRENT APPLICATION NUMBER: US/10/408, 765A
 ; CURRENT FILING DATE: 2003-04-04
 ; NUMBER OF SEQ ID NOS: 3077
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 254
 ; LENGTH: 3225
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-408-765A-254

Query Match
 Best Local Similarity 21.3%; Pred. No. 1.2e-47;
 Matches 599; Conservative 529; Mismatches 963; Indels 718; Gaps 119;

QY 251 RLKGCNRSFILGQVKKLSDQGVGGFI---NYRQSKLTRI-----L 292
 Db 680 KAKEISLNQ---LIEBPKWADNNSAFTALSBEDQLLSQVKELSMVTETRAQVKOL 735
 QY 293 QNSLGGNPKTRITCTITPVSFDETLT---ALQFASATKMTKNTPVNVEV-----STD 341
 Db 736 EMNLAEARQRLDYESTAHNLLITEQIHSLSIEAKSKVKIEVLQNELDDVQLQFSEQ 795
 QY 342 EALLKRYKPEIMDKQLQEEVSLFRAQAMEKDLQALLERKOLLQKQVONKIEINLRML 401
 Db 796 STLRSQSQONKESEVLEGAERVHHSKVSELSQALSQKEL-----EITQMDQLL 848
 QY 402 VTSSS---LTLQOELKAKRKRVTWCLGKINKMKSNDYADQFNPINITTKHLSINLIR 459
 Db 849 LEKKRDVETLQOTTEKQOQVTEISFMTKQWQVQLN---EEKFSLGVFI---KTLKEQLNLS 905
 QY 460 EIDES-----VCSSEDFVNTLDTLSEIPWNPATKLLNQNTESELNSLRADYD-----508
 Db 906 RAEAEKKEQVEDNEVSGGLQNYDEM--SPAGQ-ISKELQHFDDLKKEQKRRKLQ 962
 QY 509 -----NLVDVLEQLRTE-KEEMELKPKKNDLDEFEALERK-----TKK 546
 Db 963 AALNRKELLQVRSRLREELANLNDKSKGTEPLSETERGEVEEDK--ENKEYSEKCVTSK 1020
 QY 547 DQEMQLHISINLKNVLRVQNDLENELSSKV-----ELLREKEDQIK 592
 Db 1021 CQEIEIY-----LKQTISEKEVELQHRKDLKLEKLAABEQFQALVQKMNQTLQDKTNQID 1075

QY 593 KLOEYIDS-----OKLENIKMDLSYSLSIESIEDPKOMKQTLFDAETVAL-----635
 Db 1076 LLOAEISENQAIIOKLITSNITDAS-----DGDVALVKETVVISPCT 1118
 QY 636 -----DAKRESAFLSENLLEKEMKELATYTKQMDIQLYQSOLEAKKQKQVBLE 687
 Db 1119 GSSEHWPEDEEKLILALEKEQOLQKLOEALTSRKAILKKAQEKERHREELQKQKDDY 1178
 QY 688 KELQSAFNEITK-----LTSILDGKVP-----709
 Db 1179 NRIQEQFDEQSKENENIGDQRLQLOIQVRESIDGKLSTDDQESCSSTPGLPELFRATE 1238
 QY 710 -----KDLIC-----NLEEGKITDLQKEL 729
 Db 1239 QHHTQPVLESNCPDWPDSHEDASALQGTTSVAQIKALKEIAEKVELEKLVSTTSSEL 1298
 QY 730 NKEVEENEALREEV-----IILSELKSLPSEVERLRKEIQKSE-----ELHI 772
 Db 1299 TKKSEEVFQOQINKQGLIESIKTVSHEAFVHAESLQKLESSQLIAGLEHLRELQ 1358
 QY 773 ITSEKDKLFSEVVHKESRVQGLLEEIGKTKDDLATTSQSNYSKSTDQEFQNFKTLHMDPEQK 832
 Db 1359 KLDELQKLISKKEEDVYLSGQLE-----KEAALTKIQTETIEQE-DLIKALHTQLEMQ 1412
 QY 833 YKMWLEENEMNQEIYNLSKEAQFDSLSGALKATSELSYKTOELQEKTEV-----QERLNE 888
 Db 1413 AK---BHDERIKQLQVELCEMKQK-PEEIG-----EESRAKQIQKLOAALISRKALKE 1464
 QY 889 MEOLKEOLENRDPSLOTVEREKLITTEKLOOTLEEVKTLTOEKDDLKQLESQIERDOL 948
 Db 1465 NKSQOEELSARG---TIER-----LTKSLADVESQVSAQNKEDTVLGRLLALQERDKL 1517
 QY 949 KSDIHDTVMNIDTQOLRNALSKOHOBTINTLTKSKISEEVS---RNLMHMENTGETKD 1006
 Db 1518 ITEMDSILEN---QSLSSCESLKLAEGETEDKEKLVKIESLKSXIAEST-----1568
 QY 1007 EFQOKMVGIDKKDLEAKNTOTLTADVDKNEIIEQOKKIFSLIQEKNELQOQLESVIAEK 1066
 Db 1569 EWOEKHKELQKEYEILLOSSENVVS-----NEAERIQHVVEAVRQEQEYLYGKLSTEANK 1623
 QY 1067 EOLKTDLKENIEMTIENTOEELLLGDELKQOEIV-AQKNHAIKKE-----GELSTC--1119
 Db 1624 KETEQLOEAOBEMKEMKVRKFAK--SKQKILEBENDRLRAEVHPAGDTAKECNE 1681
 QY 1120 -----DRLAEVEEKLKEKSOLOEQKQOLLNVQSEMSEMQKINEIENKNEK--NK 1170
 Db 1682 TLLSSNASMKEELERVKMEYETLSKKFOSLMSKDSLSE-----EVQDLKHQIEGNVSK 1735
 QY 1171 ELTLEHMETERLELAQKLNENYEVKSYTKERKVLKELQKGFETERDHLRGVIRIEATG 1230
 Db 1736 QANLE--ATEKHD--NOTNVTEEGTOSIPGETE-----EODLSLSTR---PTC 1777
 QY 1231 LOTKEELKIAHILKEHQETIDELRRSVSEKTAQIINTQDLEKSHTKLOERIPVLHERQE 1290
 Db 1778 SESVPSAKSANPAVSKOPSSHDEINNYL-----QQID---QLKRIAGLEEEKQ 1823
 QY 1291 LLPNVKVSQETQETWNELELLELTQOSTTKDSTTLARIEMERLRNLEKQESQESBEKSLTKE 1350
 Db 1824 ---KNKFSQTLN--NEKNVLLSQISTDG-ELKQLOEFTVMNLLNQIIOELSRVTK-1876
 QY 1351 RDNKTTKEALEVKHDQJKEHIRETLAKIQESOSKQESQKQSLNMKEDNETTKIVSEMGEQK 1410
 Db 1877 -----LKETAEEBKDLERLNMQLAELNGSIGNYQDVTDAIQKNE-1919
 QY 1411 PKDSALLRIEIMGLSKRLQESHEMKSVAEKDDIQLRQLEVLQSESDQIKENIKEIVA 1470
 Db 1920 -----LESEMKNLKCKVSELEEEKQOLVKEKTV-----ESEIRKEVLEKIQ 1962
 QY 1471 KHLE--TEEBELKVAHCLCKEQTETINELRVNLSSEKETEISTIQKLEAINDKLQNKIOEI 1528
 Db 1963 AQEPGNKSHAKLOELLKKEQEVQKQKQDQKQIRYQEKISALERTVKAL-----EF 2013

Db 367 AVVNEDTQGNYSQOAEVRLKEQELAEASGQTPPEFLTRDKKTKNTMYEQEAMLEFFK 426
 Qy 381 -----BEKDLLOKVQNEKIENTRLMVTSSSUTLOQELKAKKRRVTCGLGKINKMKNY 436
 Db 427 KSEQEKSLIEKV-----TQLEDLTLLKKEKFI-----QSNKMIVKFR 463
 Qy 437 ADQENIPNITTKHLSI-NLLREIDSVCSDFVFNITDITLSE-TWNPATKLLNOE 494
 Db 464 EDQI-----IRLEKHKESRGGLPEQDRLLSE--LRNEITQLREQIEHPRVAKYAME 516
 Qy 495 NIESELSRADYDNLVDYQRLATEKEEMELKLEKNKNDLPEFALERKTKKQDMQLIH 554
 Db 517 N-----HSLR-----BENRLRLLE-----PVKRAQEM----- 539
 Qy 555 EISNMLVHRVYNODLENELSKVELLEKEDQIKLQYIDSQKLENIKMDLSVSL 614
 Db 540 ---DAQTIKLEKAF-----SEISG-----MEKSD--KNQGF----- 567
 Qy 615 ESIEDPKOMQTLDFAETVALDAKRESAFLRSENLEKEMK-ELATYKQEMENDIQLYQ 673
 Db 568 ---SPAQKEPCLFANT-----EKLKAQLQIQTELNNKQYE 603
 Qy 674 SOLEAKKMOVDLEKELQSAFNEITKLTSLIDGKVPKDLLCNLEEGKITDIQK-ELNKE 732
 Db 604 EFKELTEKROLESELOS-----LOKANINLE 631
 Qy 733 --VEENEALREEVILLSELKSLPSEVERLRKEIQDKSEELHIITSEKDLFSEVVKESR 790
 Db 632 NLLEATKACKRQ-----EVSQNK--IHAETIKIITP-----TKAYQLHSR 671
 Qy 791 VQGLLEEIGTKDLDATTOSNYKSTQEFQNFKTLMDPEQKYMWLE--NERMNQEI 847
 Db 672 -----PVFKLSPMGSGFSLYT-----QNSILDND-----ILNEPVPMWNEQAF 712
 Qy 848 VNLSEAKQFDSLSGLAKFISLQEKNELOQMLESVIAEKEQLKTDLENIENTIENQ 1084
 Db 713 EALSEBELTVQEQMSALQAKL---DEEHKULKQOHVDKBBHSTQMOELFS----- 762
 Qy 908 REKTLITEKLOQLEVKVLTQEKDOLKQLESQIETQERDQKSDIHD--TVANNIDTQEQ 965
 Db 763 -----SERIDWTKQOELLQNLVLEKQLOET-QTKNDFLSEVHDLRVLHSADKE-- 813
 Qy 966 LRNALESKQHOQTINTLASKISEEVS-RNLHWEENTGETKQEFQKMGVIGDKQDLEAK 1024
 Db 814 ---JSSVKLEYSFKTNOEKFNKLSERHMHV-----QLQDNLRLNEKLESK 860
 Qy 1025 NTQTLTADYKNEIIFQQRKIFSLQEKNELOQMLESVIAEKEQLKTDLENIENTIENQ 1084
 Db 861 ---ACLQDSYDN--LQEIIMKF-----EIDQLSRNLQNFKENETLKSDDLNNLMEL-LEAB 909
 Qy 1085 BELRLILGDELKQOEIVQEKKNHAIKKEGELSRTCDLRAEVEEKLKESQOQEQOQULL 1144
 Db 910 KE---RNNKLSIQFE---EDKENSCK-----EILKVLAEVQEKQKETA 947
 Qy 1145 NVQEMSMBQKINIEINLKNELKNKELTLEHMETERLELAQKLNENEVEEVKSYTKERKV 1204
 Db 948 KCEQMAKVOK-----LE-----ESLILATEKV 969
 Qy 1205 LKELOKSPETRRDLHRLGYIRIEATGLQTKBELKIAHLKHEQETIDELRSVSEKTAQ 1264
 Db 970 ISSLEKSRDSKVVADLMNQIQ-----ELRSSVCEKT-E 1003
 Qy 1265 LINT--QLEKSHYKL-----QEBIPVLHBEQELLPNVKVSEVQETMNELELITQEST 1316
 Db 1004 TIDLKQELKQDINCYNKNSALVDRESRVLIKKQEV-----DILDLKELT-LRILISE--- 1054
 Qy 1317 TKDSTTLARIEMERLRINEKQESQEBEIKSTYKERNLTKITKEALEVKHDLKHEIRETL 1376
 Db 1055 -----DIBRDMCEDLAHATEQLMMLTE-----ASKKHSGLQSAQEL 1093
 Qy 1377 AK-----IQESQSKQEQSLNMKEDNETTKYIVSEMEQPKPKDSALLRIEIMGLSKRLQE 1432
 Db 1094 TKKEALQELQHK---LNQKKE-----EVEQKKE-----YNFKMRQ 1127

Qy 1433 SHDEKSVAKKDDQLQRLQEVLOSQESDQLKENIKEIVAKHLET-BEELKVAHCLKEQBE 1491
 Db 1128 LEHYMDSAED-----PQSPKTPPHFQTHLAKLLETQOEIEDGRASKTSLKH 1175
 Qy 1492 TINELRNLSEKETEISTIOKLEAINDKLQNKQEIYEKEEQLNIKOISEVQENNELK 1551
 Db 1176 LVTKLNDREVNABILRMKEQLREM-ENLLESQOLIEKNWLL-----QGOLDIDK 1226
 Qy 1552 QFKEHRKAKDSALQSIKSMLE-LTNRLQESQEBIQIWKIEKEMKRYQEA- 1602
 Db 1227 ROKENSQNHDPNQLKNEQESIKERLAKS-KLVEEMLKMADELEVQSYKNKEMECL 1285
 Qy 1603 ---QIERDQKENTKEIVAKKESQEKYQFLQMTAVNETQEKVC-EIEHLKEQFETQK 1657
 Db 1286 RMTDEVERTQTLSE-----KAFQEKQELRSKLEEMYEBERTSQEMEMLRKQ----- 1332
 Qy 1658 LNLNIEETENIRLT--OILHENLEEMSVTKERDDLSRVEETLKVRRDQKLENLR 1710
 Db 1333 --VECLAENGKLVGHQNLHQKIQYVVRLLKENVRLEAETEKLAENVFLKEKR 1385

RESULT 5
 US-10-146-473-82
 ; Sequence 82, Application US/10146473
 ; Publication No. US2003010888A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Scanlan, Matthew
 ; APPLICANT: Gout, Ivan
 ; APPLICANT: Stockert, Elisabeth
 ; APPLICANT: Gure, Ali
 ; APPLICANT: Chen, Yao-Tseng
 ; APPLICANT: Old, Lloyd
 ; TITLE OF INVENTION: Breast Cancer Antigens
 ; FILE REFERENCE: L00461/70130(JRV)
 ; CURRENT APPLICATION NUMBER: US/10/146,473
 ; CURRENT FILING DATE: 2002-05-15
 ; PRIOR APPLICATION NUMBER: US 60/291,150
 ; PRIOR FILING DATE: 2001-05-15
 ; NUMBER OF SEQ ID NOS: 82
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 82
 ; LENGTH: 1388
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-146-473-82

Query Match 8.1%; Score 1077.5; DB 14; Length 1388;
 Best Local Similarity 26.5%; Pred. No. 1.1e-45;
 Matches 475; Conservative 280; Mismatches 521; Indels 519; Gaps 79;

Qy 3 BEGAVAVCVVRPLNSREESL-GETAQVYWKTDNNVIYQVDGS-----QNLCLSVLSSTLRHSNPEPKTFD 73
 Db 23 EGDALKVFRIRPPAERSGSADGE-----QNLCLSVLSSTLRHSNPEPKTFD 73
 Qy 51 RVFHGNETTKNYVEIAPIIDSALQNGTIFAYGQTASGKTYTMMG-----SEDLH 103
 Db 74 HVADVDTTQESVFATVAKSIVSCMSGYNGTIFAYGQTGSGKTTMMGESDSENFHNL 133
 Qy 104 GVIPRAHDIF-----OKIKKFPDREPLLRYSYMEIYNETITDILCGTQKMKPLIREDV 158
 Db 134 GVIPRSEYLSLDRKEKAGAKGKFLCKSFIEIYNEQIYDLL--DSASAGLYRSHI 191
 Qy 159 NNNVYADLTVEVVYVSEMALKWITKGEKSHYGETKMNQSSRSSTIFRMLLSREKGE 218
 Db 192 KKGVVVGAQVQVTSAAAYQVLSGGWRNRVASTMNRSSSHAVFTIISMEK-- 249
 Qy 219 PSNCEGSVKSHMLVDLAGSRAAQTGAAGVRLKEGCNINRSIFILQVILKLSGQGV 278
 Db 250 -SNEIVNIRTSLLMLVDLAGSERQKTHABGMRLKEAGNINRSLSCLGQVITALVD--VG 306
 Qy 279 G-----EINYRDSKLTRLQNSLGNPKTRICITIPVS--FDETILTALQFASHTAKYKMT 332

```

307 NGQHVYCDYRDKLITLLRSLGNAKTAIIANVHFGSRFCGETLTLNFAQRAKLKNN 366
333 PYNE-----VSTDEALLKRYKEIMDLKK-OLEEVSLETR-----AAMEKQOLAQL-- 380
367 AVUNEDTQGNVSQLOAEVRLKEQLAEASGOTPPESFLTRDKKKNYMEYFOEAMLFFK 426
381 -----BEKDLLOKVQNEKIENITMLVTSSSLTLQELKAARKRVWCLGKINMKNSNY 436
427 KSPQEKSLTEKV-----TOLEDTLKEKFI-----QSNKMIVKPR 463
437 ADOFNIPNTITTKHLSI-NLAREIDSVCSRSDFVSNLTDLTSE-IENWPAKLNLQOE 494
464 EDQI-----IRLEKLHESRGFLPEEQDRLSE--LRNEIQTILREQIEHHPVAKYAME 516
495 NIBSELNSLRADYDNLVDYEQLETEBEMELKLKEDNDLDEFALERTKTKDQEQMLIH 554
517 N-----HSLR-----BENRRLLE-----PVRAQEM----- 539
555 EISNLKVLHREVVYNQDLENELSELKVELLREKEDQIKKLOEYIDSOQLENIKMDLSYSL 614
540 ---DAQTAKLEKAF-----SEISG-----NEKSD--KNQOGF----- 567
615 ESIEDPKOMQTLFADAEVALDAKRESAPLSENLKERNK-ELATTYKOMENDIQLYQ 673
568 ---SPKAQKEPCLFANT-----EKLKAQLLQIQTLENNKSKOYE 603
674 SOLBAKKQMDVLEKELOSAFNEITKLTSLIDGKVPKDLLCNLEGGKITDLOK-ELNKE 732
604 EFKELTRROLESELOS-----LOKANLNL 631
733 --VENEALREVILLSELKSLPSEVERLKEIOPKSEBELHIITSEKDLFSEVVHKESR 790
632 NLLKATKACKQ-----EVSQNLK--IHAETLKLIITP-----TKAYQLHSR 671
791 VQGLLEIEGKTDDIATTQSNYSTDQFONPKTLHMFQEQKVMLEB--NERMNQEI 847
672 ---PVPKLSPEMGSGSLYT-----QNSILDND-----ILNEPVPEWNEQAF 712
848 VNLSEBAKFDSSGALKATELYKQLOEKTREYQERLNEQKLEQLENRDSPLQTV 907
713 EAISEBELRTVQPMGALQAKL-----DEEHNKMLKQOHVDKLEHISTQMFELS----- 762
908 REKTLITEKLOTEVEVTLTKQDKQLQBSLQTERDOLKSDIHD--TVANNIDTQOQ 965
763 -----SERIDWKQOEELLSQLNVLEKQLOET-QTKNDFLKSEVHDLRWLHLSADKE-- 813
966 LRNALESKQHOETINTLKSKEEVS-RNLHMEENTGETKDFEQKMGVLDKQDLEAK 1024
814 ----LSSVKLEYSSFKTNOBKFNKLSERHMHV-----QLQDLNLRLENEKLESK 860
1025 NTQTLTADVKDNEIIBQOKIFSLIOEKNELOQMLSEVIAEKEPOLKTDLKENIEMTIENQ 1084
861 ---ACLQDSYDN--LOEIMKF-----EIDQLSNQLNFKKENETLKSDDLNNLMEL-LEAE 909
1085 EELRLLGDELKQOETVAOENKHAIEKEGELSRTCDRLAEVEKLEKESQOQKQOQLL 1144
910 KE--RNNKLSLOF-----EDKENSCK-----EILKVLAEVROEKOKETA 947
1145 NVQEMSEMOKKINEIENLNKELNKNKELTLEHMETELELAQKLNENVEVKSITKERV 1204
948 KCEQMAKVQK-----LE-----ESLLATEKV 969
1205 LKELOKSFEDRHLRGYIREIATGLQKELKIAHILKEHOETIDELRRSVSEKTAQ 1264
970 ISSLEKSRDSDKKVVADLMNQIQ-----ELRTSVCEKT-E 1003
1265 IINT--ODLEKSHTKL-----QBEIPVLHBEQBLPNVKVSTQETMMELELLTQOST 1316
1004 TIDTLKQELKDJNCKYNSALVDEESVLLKKQEV-----DILDKETL-RLRLISE-- 1054
1317 TKDSTTLARIEMERLRNKEQOEQBEIKSTKERDNLTKITKEALEVKHDKQLEHIRETL 1376
1055 -----DIERDMLCEDLAHAETQNLNTE-----ASKHSGLLQAQEL 1093

```

RESULT 6

```

US-10-173-999-32
; Sequence 32, Application US/10173999
; Publication No. US20040005563A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Ovarian Cancer, Compositions
; TITLE OF INVENTION: and Methods of Screening for Modulators of Ovarian
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 018501-002420US
; CURRENT APPLICATION NUMBER: US/10/173, 999
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/299,234
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/315,287
; PRIOR FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 163
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-173-999-32

```

Query Match 8.1%; Score 1077.5; DB 15; Length 1388;
 Best Local Similarity 26.5%; Pred. No. 1,le-45;
 Matches 475; Conservative 280; Mismatches 521; Indels 519; Gaps 79;

```

QY 3 EGAIVAVVRVPLNREESL-GETAQVYVKTNNVYQVDS-----KSFNFD 50
DB 23 EGDAIKVVRIRPPAESGSGADGE-----QNLCLSVLSSTLSRLHSNPPKPTFFD 73
QY 51 RVFHGNETTKNVVEETAAPIIDSAIOYNGTIFAYGQTASGKYTWG-----SEDHL 103
DB 74 HVADVDVTTQBSVFATVAKSIVSCMGYNGTIFAYGQTGSKGKFTWMPGSDNFNHLR 133
QY 104 GVTPRAIHDIF-----QKIKKFPDRFELLRVSVMEIYNETITDLLCGTQKMKPLIREDV 158
DB 134 GVTPRSEYFLSLIDREKEKAGAKSFLCKSFIEIYNEQIYDLL--DSASAGLYLREHI 191
QY 159 NRVYVADLTVEVYVSEMALKWITTKGKSHYGETQNNQSSRSHITFRMILESREKGE 218

```


23 EGDAIKVFRIRPPAERSGADGE-----QNLCLSLVSTSLRLHNSPCKPTFTD 73
51 RVPHGNETNVTVEETAAPIIDSAIOYNGTIFAYGOTASGKTYTWMG-----SEDL 103
74 HVADVDTTOESFATVAKSIVSCMGYNGTIFAYGOTSGKTYTWMGSESNFHNLR 133
104 GVTPRAHIDF-----QKIKFPDRFLRVSMEIYNETITDLCGOTKMKPLIREDV 158
134 GVTPRFEYLFSLIDREKAGAGKAGKFLCKSCFIEIYNEQIYDL--DSASAGLYREHI 191
159 NRVVYVADLPEEVVYVSEMAKWTIKGKSRHVGTEGONORSRSHITFIMILESKEGE 218
192 KKGFFVVGVEQVVTSAEAYQVLSGCMWRNRVASTSMNPRESSHAFTITTESMBK-- 249
219 PSNCEGKVKSHNLVLDLAGSRAAQTGAAGVRLKEGCNINRSILFGLQVKKLSDQVG 278
250 -SNEIVNIRTSLLNVLVLAGSEROKTHAEGMRUKEAGNINRSLSCLGVITALVD--VG 306
279 G-----FNVRDSKLTILONSLGPNKTRITCTTPVS--FDETLTALOFASTAKYMN 332
307 NGQRHVICYRDSKLTFLRDSLGNAKTALIANVHPGSRGCFGETLSTLNFPAQAKLIK 366
333 PYNE-----VSTDALKRYKREIMDKK-OLEEVSLETR-----AQAMEKOLAQL-- 380
367 AVNEDTQGNVSQLOAEVRLKEQALASQTPPESTFLTRDKKTYMEYFOEAMLFFK 426
381 ----EEDOLLQVONEKIENITMLVTSSSLTLOELKAKKRRVTCGLGINKMKNSNY 436
427 KSEQKKSLEBK-----TOLEDLTLEKKEFT-----OSNMIVKFR 463
437 ADOFNIPNITTKHLSI-NLLREIDSVCSSESVFNTSLTDLSE-TEWNPATKLINQE 494
464 EDQI-----IRLEKHKSRGFLPEEQDLASE--LNEIQTLEQIEHPRVAKYAME 516
495 NIESELNSTRADYDNLVDYQLRTEKEMEELKKEKNDLDEFALERKTKKQEMQLIH 554
517 N-----HSLR-----EENRRLLE-----PVKRAQEM-- 539
555 EISNLKNLVKREYVNOLENELSKVLELLREKEDQIKKQBYDSQKLENKMDLSYL 614
540 ---DAQTTAKLEKAF-----SEISG--MEKSD--KNOQF----- 567
615 ESIEDPKOMKOTLFDATVALDAKESAFRLSENLEKEMK-ELATTYKOWENDIQLVQ 673
568 ----SPKAQKEPCLFANT-----EKLKAQLLIQIOTELNNSQBYE 603
674 SOLEAKKKQVQVLEKELQSAFNEITKLTSLIDGKVPKOLLNLELEKIDTDLQ-ELNKE 732
604 EFKELTRKQLESELSQ-----LOKANINLE 631
733 --VEENEALREEVILLSELKSLPSEVERLRKEIQDKSEELHITSEKDLFSEVVKESR 790
632 NLEATKACKRQ-----EVSQNK--IHAETIKLITP-----TKAYOLHSR 671
791 VQILLEIGTKDDLATTOSYKSTQDFQNFKTHMDPEQKYMWLEE--NERWNQBI 847
672 ----PVKLSPEMSGFSGLYT-----QNSILND-----ILNEPVPPEMNEQAF 712
848 VNLSKEAQKPDSSLGALKTSLVKTQBLQKTEVREQLNEMQLEKLENRDSPLQTV 907
713 EASEELRTVQEQMSALQAKL-----DEEHNKLOOHVDKLBHSTQMELFS----- 762
908 REKTLTEKLOOTLEEVKTLTOEKDDLQOLQESLQTERDQKSDIHD--TVNNNIDTQEQ 965
763 ----SERIDWTQKQBELLSQLNVLEKQLOFT-OTKNDFLKSEVHDLRVVLHSADK-- 813
966 LRNALESKLOHETINTLKSISEVS-RNLHMEENTGETKDFQKQWGDKKQDLK 1024
814 ----LSSVKLEYSFKTKQEKFNKLSERHMHV-----QLQDLNLENEKLESK 860
1025 NTOTLTADVKNDNEIIOQRKIFSLIOEKNELOQMLESVIAEKQKTDLENEMTENTQ 1084
861 ---ACLOQSDYN--LOEIMKF-----RIDQLSRNLQNFKKENETLKSOLNLMEL-LEAE 909

1085 BELRLGDELLKQOEIVAOEKHAIKKGELSNTCDRLAEVEBEKLEKESQOQLEKQOQL 1144
910 KE---RNNKLSLOFE---EDKENSX-----EILKVLAVRQEKQKETA 947
1145 NVQEESEMQKINEIENLNKLNKELTLEHMETERLELAQKLNENYEVKSIKERV 1204
948 KEOQMAKQVK-----LE-----ESLLATEKV 969
1205 LKELQKSFTEHDRLRGYIREIATGLQYKEELKIAHILHKEHOETIDELRRSVSEKTAQ 1264
970 ISSLESRSDKKVADLMMNQI-----ELTSCVCEK-E 1003
1265 IINT--ODLESKSTKL-----OEETPVLHEQBELLPNVKVSETOFTMNELELLTQST 1316
1004 TIDTLQELKDINCKNSALVDRSESVLKKQEV-----DILDKETT-LRLILSE-- 1054
1317 TKDSTTLARIMERLRINEXFQSEBEIKSLTKERDNLKTIKEALEVKKHOLKHEIRETL 1376
1055 -----DIERDMLCEDLAHATEQLNMLTE-----ASKKHSGLQSAQBEL 1093
1377 AK-----TOESQSQOEQSLNMKEDNETTKIVSEMOEFPKDSALLRIEIMLGLSKLOE 1432
1094 TKREALQELQHK-----LNQKE-----EVEQKNE-----YNFKMRQ 1127
1433 SHDEMKSVAKEKDDLQRLQEVLOSQDLKENIKEIVAKHLET--EELKVHCCLEKQOE 1491
1128 LEHYMSAAED-----POSPTPHFQTHLAKULETOEQEIEDGRASKTSLEH 1175
1492 TINELRVNLSKETEISTIQOLBAINDKQNKIOEYKKEEQLNIKOISEVOENNELK 1551
1176 LVTKLNEDEVKNAELRMKEQLREM-ENLRLESQOLIEKNWLL-----OQQLDDIK 1226
1552 QFKERKAKDSALOSIESKMLE-ITNRLQESQOEIQTIMKEEMKRVQOEAL----- 1602
1227 ROKENSQDHPDNOQLKNEQESIKERLAKS-KIVEEMLKWKADLEEVQSALYNKMECL 1285
1603 ----QIERDOLKENTKEIVAKMKESQEKYQFLKMTAVNETQKMC-BIEHLEKQEFETQK 1657
1286 RMTDEVERTOTLES-----KAFOEKEQLSKLEEMVEBERRTSQEMELRKQ----- 1332
1658 LNLNIETENIRLT--QILHENLEEMRSVTKERDDLRSDRSVEETLTKVERDOLKENLR 1710
1333 --VECLAENGKLVGHQNLHQIQYVVRKLKENVRLABETEKLRANVFLKEKR 1385

RESULT 8
US-10-171-311-4
; Sequence 4, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerish, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: Fast-Seq for Windows Version 4.0

```

; SEQ ID NO 4
; LENGTH: 3899
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-4

Query Match      8.0%; Score 1064; DB 14; Length 3899;
Best Local Similarity 19.7%; Pred. No. 1.9e-44;
Matches 672; Conservative 618; Mismatches 1107; Indels 1010; Gaps 150;

QY 26 TAQVWYKTDNNVYOVYDYSKSFNDRVPHGN-----ETTKVYBIAPIIDSAIQYNGT 81
DQ 295 TMOISFLQEKIVYEMQDK-----KVNSNKBEIQKETIIBELNATIIIE----- 341
QY 82 IFAYGQTASGTYTMMG-----SEDLHGVIPRAIHDFQIKTKPPDREFFLRVSMELYN- 136
DQ 342 -----EKKTELEKDKLTADKLLGELQEQIVQKNQEIKNMK-----LEITNS 383
QY 137 -----ETITDILCGTQMKPLIITREDVNRVYVADLFEVYVYTSSEMALKWITTKGS 188
DQ 384 KOKERQSSEIKQMGTVIELQ-----KRNHKSQFETDIVORMEQETQKLEQL-RAELD 438
QY 189 RHYGE--TKNQSSSSHTIFRMILSRKGEPSNC-----EGSVKVSHLMDVL 236
DQ 439 EMYGQIVQMKQELIRQHWAEEMKTRHKGEMENALRSYNTVNEDQIKLMVAINEL 498
QY 237 -----AGSERAAQTGAAGVRLKEGNCINRSLFILGVITKLSGDG-----OVGGINYRD 285
DQ 499 NIKLQDNTSOKKEKLEELGILKEKCALQKQ-----LEDVLELSFSREQIQRARQTTABQE 555
QY 286 SKLTRILQNSLGNPKTRIICITP-----VSPDETALQFASHTAKYMKNTPYVNEVST 340
DQ 556 SKL-----NEAHKSLSTVEDLKABIVSASESRKELE-----LKHEAEVINYKI 598
QY 341 DEALKRYRKEIMDLKQLEEVSLR-----TRAQAMEKDQLAQLLEKDLQKQVNEKIENL 397
DQ 599 KLEMLEKKNVLDMAESQBAELRLTQLLPSHEEELSKLKDLEIEHRINIEK----- 654
QY 398 TRMLVTSSSLTIQOELKAKRRVTCWLGKINKMKNVYADQFNPTNITTTTKHK-----LSI 455
DQ 655 -----LKNLGIHYKQID--GLQNMESQKIETMQFE-KDNLITKQNLILIEI 699
QY 456 NLLREIDES-VCSSEVDSNTLDTL-SEIENPDKLQENIRSELNSLRADYDNLVLD 513
DQ 700 SKLDLQOQSLVNSSEMTLQINELQKETE-----ILRQE--EKEKGTLEQEVQEL-- 748
QY 514 YEQLRTEKEEMELKLEKNDLDE-----536
DQ 749 --QLKTELLEKQKKE-NDLQEKFAQLEAENSILKDEKKTLEDMLKIHTPVSQERLIF 805
QY 537 PEALERKTK-----KQEMQLIHEISNLKNLV-----KHREVNQDLE 574
DQ 806 LDSIKSKSDSWEKEIEILTEENEDLQOQIQLNEEIEKQENTSPAKKFEVNYQELQ 865
QY 575 NELSSKVELLREKEDIKIQEYIDSQKLEIKMDLSYLSIESIEDPK-OMKQTLFD----- 629
DQ 866 EYACLLKVKODLEDS-KNKQLEYKSKLKALNEEL--HLQRIINPTTKMSSVFDEDKT 922
QY 630 --AETVAL-----DAKESAFLRSENLEKMKELATTYKQENDIQLOSQLEA 678
DQ 923 FVAETLEGEVVEKDTTLEMKLEVTREKLELSQRLSDLEQKQKHGHSFLNEEYKS 982
QY 679 KK--KMQVDLE-KELQSAFNE-----ITKLTSLIDGKVPKDLLCNLEKGTIDQLQ 726
DQ 983 LKQEKQVSLRCRELEIINHNRAENVQSCDTQVSSLLDGVV-----TWTSRGAESVSQVN 1039
QY 727 KELNKE-----VEE-----NEALREEVILLSELKSL-----PSEBRLRKEI 763
DQ 1040 KSGFEESKIMVEDKVSFENMTVGTESKQEQILDLPLSPVTKESSILRATOPSENDKQLQEL 1099
QY 764 QD-KSEE-----LHI-----ITSEKDK-----LPSVWH-KESRVQ 792
DQ 1100 NVLASEQNDLQWAEQICLSLVYTHVDQVRYMEYNEKDKALCSLKEELIFAQEBKIK 1159

```

1614 KEIVAKESQOEKEYOFLKMTAVNETQKMCIEHLKEQFETOK-----LNLENIENTIR 1669
2195 KEIT-----NLEBOLEQFRE-----ELENKNEEVOQLHMQLIEIQKKESTTFLQOLEQEN-- 2243
1670 LTOILHENLEEMSVTKERDILASVEETL-----KYERDOLKENL----- 1709
2244 --KLPDDMEKGLAIKESDAMSTQDQHVLFGRFAQIIQOEVEIDQLNEQVTKLOOQLK 2301
1710 -----RETTIDLEKQOEELKIVHML--KEHQETIDKLRGIVSEKTNISNM-OK 1756
2302 ITTNDKNVIEKNELIRDETQIECLMSDQECVRNREEIEQLNEVIEKLQOELANIGQK 2361
1757 -----DLESNDALKAQDLKIOELRIAH----- 1780
2362 TSMNAHLSSEASLSKHQLOVVIKALALQOQVETANEEMTPMKVNLKTFNFKMNQLTQE 2421
1781 --MHLKEQOETIDKLRGIVSEKTD--KLSNMOKDLENSNAKIOEKIOELKANEHQ----- 1831
2422 LFSIKRERESVEKIQSPENSVMVAIDHLSKDKPELEWLTE--DALKSLENQTYPKSFE 2479
1832 -----LITLKOVNETOKKVS-----EMEQLKKQIKDOQS-----LTLKLE 1867
2480 ENKGSIINLETRILLOLESTVSANDLELTQYKIQKDMQBGQOPETEMLOKIVNLQKIV 2539
1868 IENLNLAEQELHENLEEMKSVMK-----ERNLRRVEETLKLDRDQKESLOET 1915
2540 EKVAAALVSIQIEAEOVYAKFQDNQTTISSEPERINIQNLQ--LREDELGSDISAL 2596
1916 KARDLEIQO---ELKTARMLSEKHEKTVDKLREKISEKTIQISDIQKDLKSKDELQK 1972
2597 TLRISELESQVVMHTSLILEKEQVEIAEK--NVLEKEKKLQELKLEGNKKORKE 2653
1973 QELQKKEQLQLRVEDVNMHKK--INEMOLKQKQEPNYLCKEEMDNFOLTKKLHESL 2029
2654 KRSQDDVEVLKTTTELPHSNEESGFNELEALRAE---SVATKAEIASY--KEKAEL 2707
2030 EEIRIVAKE-----RDELRIKESLK--MERDQFIATLREMIARDRONHVKP 2075
2708 QE--ELVVKETNMTSLQKDSQVRDHLAEKELKISLEKETEVEQESKACWFEPILKL 2766
2076 EKRLIS--DGOQHLMES-----LREKCS--RIKELLKRYSEMDDHYECL 2115
2767 SKSIASQTDGTLKISSNOTPOLVKNAGIQINLQSECSSEVETIISQFTEKIEKMQEL 2826
2116 NRLS--LDLE--KEIEPHRMKLVLYSVYTKIEB-----QHECINKFEMDFIDEV-- 2165
2827 HAAEILDMESRHISSETLTKREHYA--VOLKEECGLTKAVIQLRSKEGSSIFELAHS 2884
2166 --KOKELL-----IKIQLQDCDVPES--RELRDLKLNQNMDLHIEILKQFS 2209
2885 DAYQTRICSSDSGSDWGQIYLTH--SQGFDIASGRGESESATDSPFKIKGLLR--- 2940
2210 ESEFPSIKTEFOQVLS-----NRKEMTOFLEWLNTR-----FDIEKLKNGLOK--EN 2255
2941 -----AVHNEGQVLSLTPSYSDGSDHDSIQVSEFWLEERKAYINTISSLKLITRMQL 2995
2256 DRICQVNN-----FFNNRIIATMNESTE----- 2278
2996 QREAEFYDSSQSHESDMRGELLALQVFLERSVLLAAPTETALTGTDTDAVGLLNC 3055
2279 FEERSATISKWEQDLSLKEKNEK--LFKNYQTLKTSLSAGAQVNTTQDNKP----- 2331
3056 LEQRIQOQVEYQAAMECLQADRRSLSEIQAALHAQMGNGRKITLREQOESKPSOELLE 3115
2332 -----HVTISAT-----OLITEK--IRELENSLHEAK-----ESAMH 2361
3116 YNIQOQSOMLEQVLSMKDRATELOELSSKRMVVAELKSELACTKLELTTLKAQH 3175
2362 KESKIKMQK-----ELEVTNDILAKLQAKVHESNCKLEKTKETIQVLDQKVALGAK 2413
3176 KHLELFAFRLEVKDQTDVHLLNDTLASEQKSLRELQWALEKEKA-----KLGRSEE 3228
2414 PYKEEIEDLKMVLKVIDLEKMKNAKEFEKISATKATVEYQKEVIRLLRENLRSSQAQD 2473

3229 RKHEELEDKPSL---ESQKQNLQ-----LNLLLEQKQLLNESSQKIESQRMLYD 3277
2474 TSVISHTDPOPSNKLTCGGSGIVONTKALILKSEHILKEIKLEKIQKQONE----- 2526
3278 AOLSEE---QGRNLEL-----QVLLSEKVRIREMSSTLDRERELHAQLOS 3320
2527 -----OLIKQKNELLNNOHLSNEVKTWKERTLK-----REAHK 2560
3321 SDGTGOSRPPSPEDLKLQKLEBKHSRIVELLNETEKYKLDLSLQTRQOMEKDRQVHR 3380
2561 QVTCENSPKSPKVTGTASKKQITPQCKERNLQDPVPKESPKSCFPDSKSLSPSPHPV 2620
3381 KTL-----QTEQANTGQKK--MHELQSKVEDLQRLQEEKQVYKLDLEGQRLQ----- 3429
2621 RYFDNSLGLCPVQVQNAESVD-----SOPGFWHASSGK 2655
3430 -----GIMQFQKQELEREBKRESRRILYQNLAEPTTWSLTSR 3468

RESULT 9
US-10-171-311-8
; Sequence 8, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Ganuvarapu, Manjula
; APPLICANT: Hoersb, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 3917
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-171-311-8

Query Match 8.0%; Score 1064; DB 14; Length 3917;
Best Local Similarity 19.7%; Pred. No. 1.9e-44;
Matches 672; Conservative 618; Mismatches 1107; Indels 1010; Gaps 150;

QY 26 TAQVYKMTDNVYQVDSKSFNFRVPHGN---ETTKVYVERIAAPIDSALQYNGT 81
DB 295 TWQISFLQEKIKVYEMEQDK---KVENSKEEIOEKETIIEELNTKLIEE----- 341
QY 82 IFAYGQTASGKTYTWMG---SEDLGLVPIRAIHDFQIKKFPDPREFLLRVSYMBIYN- 136
DB 342 -----EKKTLELKDXTADKLLGELQELVQKNQEIKNMK-----LELTNS 383
QY 137 -----ETITDLCGTQMKPLIIREDNVNNVYVADLTETEVVYTSMAKWKITKGEKS 188
DB 384 KQKERSSEEEKLGMGTVEELQ---KRHKDSQFETDIVORMEQETQKLEQL--RAELD 438
QY 189 RHYGE--TKMNQSSSHIFRMLLESREKGEPSNC-----EGSVKVSHLNLVDL 236
DB 439 EMTYGOQIVQKQELIRQHMARQMEEMKTRHKHGENALRSYNTVNEDQIKLMVAINEL 498

Db 806 LSIKSKSDSVWEKTEIIEENEDLKQOCIOJNBEIEKORNTFSPAENFEVYQELQ 865
 Qy 575 NEUSSKVELLEREDQIKLQEIYDSQKLENKMDLSVLSIESIEDPK-QMKOTLFD--- 629
 Db 866 EYACLLKVVDDLEDS-KNQOELEYKSKLXALNEEL--HLQRINPTTVKMKSSVFDDEKT 922
 Qy 630 --AETVAL-----DAKRESAFLRSENLEKERMKEIATTVYQWENDIOLYQSOLEA 678
 Db 923 FVAETLEMGVWEKDTTLMELKLEVTREKLELSQRLSDLSQKQKHGEISFLINEBVS 982
 Qy 679 KK--KMQVDLE--KELQSAFNE-----ITKLTSLIDGVPKDLCLNLEBEGITDLO 726
 Db 983 LKQEKQVSLRCELEIIIIHNRAENVQSCDTQVSSLLDGW---TMTSRGAESVSKN 1039
 Qy 727 KEIANKE-----VEE-----NEALREYVILLSKSL-----PSEVERLURKEI 763
 Db 1040 KSGEESKIMVEDKVSFENMTVGEESKQEOIILDLPSVTYKSSLRATQFSEMDKQKEL 1099
 Qy 764 QD-KSEE-----LHI-----ITSEKDK-----LFSVVH-KESRVQ 792
 Db 1100 NVLKSPQNDRLQMEAQRICLSIVYTHVDQVREYMEKOKKALCSLKEBILFAQEEKIK 1159
 Qy 793 GL-----LE-----EICKTKDODLATQSNYKST----- 815
 Db 1160 ELQKHOLELQTMKTQETGDEGKPLHLILGKQKAVSEBCSYFLQTLCSVLGEYTPALK 1219
 Qy 816 -----DOEPQNFKTLHMDPEQKYKVMLEB-NERMNOEIV----- 848
 Db 1220 CEVNAEDKENSVDYISENEDPELODYRYEVQDQENWHLLNKVTEYNKLLVLQTRLSK 1279
 Qy 849 -----NLSKEAQKF-----DSSLGAL-----KTELSUK 871
 Db 1280 INGQOTDGMKLBFGREENLPKEETEFLSIHSQMTNLEDIDVNHKSLSLQDLKTKLEQ 1339
 Qy 872 TQLOEKTRVORLNEMLQKLENRDPSPLQTVREKTLITEKLOQTLFEVKTUTQEK 931
 Db 1340 VOELSLSSLOQLKETEQ-----NYEAHICLQKRLQAVSESTVPPSPDPVSVVITE 1393
 Qy 932 DDLQLOESLQIERDQKSDIHDVTNMNTDTQOLRNALESKHOETINTLKSXI----- 987
 Db 1394 SDAQRTMYPGSCVKNIDGTIFSGFGEYKETEINIVKLE--KOYQOELEEVAKVIVSM 1451
 Qy 988 -----SEVSNLHWEENTGTQDEF-----QQRWVGIDKQDLRAKNTQTL-TADVK-- 1034
 Db 1452 STAFQAQOTELSRISGKENTASSKQAHAVCCQEQHYFENMKLSQDQIGQTFETVDVVEK 1511
 Qy 1035 -----DNEILBQQRKIFSLIOEKNELQOMLESVIAEKEQL-----KTDLENTEM 1079
 Db 1512 BEFKPLSELGHBGHEIILNSDPHDIPESKDCVLTISEMFSKDTFIVRQSIHDEISV 1571
 Qy 1080 TIENQOEELRLGDELK-KQOIVAQEKNAHAKKEGEL-----SRTCORLAEVEEKLAKEK 1132
 Db 1572 SSMDASRQMLNNEEQLEDWRLVQYQEH--QOATELLRQAHMROMERQREDQOLQEE 1629
 Qy 1133 SQLOEKQOQLNVQEE--MSEMQK-KINEIENLKN-----ELKN----- 1169
 Db 1630 IKRLNRQLAQRSSIDNENLVSERERVLLEELKQLSLAGREKLCCELNRSSTQNGN 1689
 Qy 1170 -----KELTLEHMETEL--ELAQKL--NENYEVKSITKERVKLQKQSF-----ETE 1215
 Db 1690 ENQGEVEQTFKEKELDRPDPVPPILSNRYALQKANNRLKYLLEVVKTAAVETI 1749
 Qy 1216 RDHLRGYI-----RETEATGL--QTEELKIAHILKHEQETIDELRRSVSEKTAQINTQ 1269
 Db 1750 GRHVILDRSSKSSQSSASLWRSBAASVSKVCHVEETRVDE---SIPSYSGDMPRN 1806
 Qy 1270 DLE-----KSHTKLOBEI-----PVLHBEQELLPNVKVSQTQTMNE--LELLITEQ 1314
 Db 1807 DINMWSKVTEBGTSLQRLVRSFAGTETIDPENBELMLNIS--SRLQAAVEKLELAISET 1864
 Qy 1315 STTKDSTTLARIEMERLRLNEKFOESQOEIKSILKERONLKIIEALEVXHDQKHEIRE 1374
 Db 1865 SSQLEHAKVTOTELMR-----ESFRQKQEATESL-----KQOELE 1901

Qy 1375 TLAKIQESQKQSQSLNMKEDNETTKIVSEM-----EQFKPKDSALLRIEIMGLSKRL 1430
 Db 1902 RLH--ESRARBEQIVALSELSKAEGVIDGYADEKTLFIERQIQEKTIDIIRLEQELLCSNRL 1959
 Qy 1431 QESHDEMSVAKKDDLQRLQOEVLQSES-----DQLEKENIKEIVAKHLET----- 1475
 Db 1960 QELEAQOQIQEBRELLSROKBAKAGPVEOOLIQETEKIMKBKLEVOQCAEKVRRDDL 2019
 Qy 1476 EEBLVAHCCLEQBQETINELRVNLSEKETEISTIQKLEAINDKLQNK----- 1524
 Db 2020 QKQKALEIDVBERQVSRFIELE--QEKNTLMDLRQONQALQKLEKMKFELDEQAIDR 2076
 Qy 1525 -----IQEIIYKEBQLNI-----KOISEVQENNELKQFKEHRKAKASALQSIESKMLE 1573
 Db 2077 EHERDVFOEQIKLEOQLKVVPRFQPISEHQ--TREVEOLANHLKEKTKDCSELLLSKEQ 2134
 Qy 1574 LTNRLQESQBEI-----QIMIKEKEEMKRYQE-----ALOIE 1605
 Db 2135 LQDIOERNEETEKLFRVRELEQALLVVSADTQFKVEDRKHFGAVEAKPELSLEVOLOAE 2194
 Qy 1606 RQOLKENTKEIVAKMKESQEKYQFLKMTAVNETQEKMCIEIHLKEQFETQK-----LNLE 1661
 Db 2195 RQADIRKEKEIT-----NLSEQLQEPFE-----ELEKNNEEVQQLHMQLEIQKKESTRLO 2245
 Qy 1662 NIETENIRLTOILHENLEEMRSVTHERDRLRSVEETL-----KVERDQLKENL 1709
 Db 2246 ELEQEN-----KLFKODMEKLAGLAIKESDAMSTQDQHVLFQKFAIIIOEKEVEIDQLNEQV 2301
 Qy 1710 -----RETITRDLKOEELKIVHML--KEHQETIDKLRGIVSEKTN 1749
 Db 2302 TKLQOQLKITDNKVIKKEKELIRDLQTEICLMSQECVKNREBIEIQLNEVIEKLOQ 2361
 Qy 1750 EISNM-QK-----DLHNSDALKAQDLKIOBELIAH----- 1780
 Db 2362 ELANIGKQTSMAVHSLSEADSLKHQDVLVIAEKLALEQOVETANEEMTFMKNVLKETNP 2421
 Qy 1781 -----MHLKEQOETIDKLRGIVSEKTD--KLSNMQKDLNSNAKLEKIQLKANERH 1830
 Db 2422 KANQLTOELFSLKRRESVEKIQSPENSNNVAIDHLSKDKPELEVLTLE--DALKSLEN 2479
 Qy 1831 Q-----LITLKKOVNETQKVS-----EMEQLKKQIKQOS----- 1860
 Db 2480 QYFKSPENGKGSINLETRLLQLESTVSAKDLQTCYKIQKOMQEQOQFETEMLQKK 2539
 Qy 1861 -LTLKSLBIENLQAELHENLEEMKSVMK-----ERDNLRRVEETLKLERDQ 1907
 Db 2540 IVNLQKIVEEKVAAALVSOIQLEAVQYAKFCODNQTTISSPERTNIQNLQ--LREDE 2596
 Qy 1908 LKESLOETKARDLEIQ--ELKTARMLSKKHETVVDKLRKISIKTIQISDIQKDLK 1964
 Db 2597 LGSDISALTIRISELESQVEMHTSLILEKEQVEIAEK--NVLEKKEKLLLEQLKLEGN 2653
 Qy 1965 KDELQKIQELQKLEQLLRVKEDVNMVSHK--INEMEQKKQEPNYLKCENMDNQL 2021
 Db 2654 EKKQREKKEKKSPODEVVLKTTTTLFHSNBSGFFNELEALRAE--SVATKAELASY-- 2708
 Qy 2022 TKQLHESLEEIRIVAKE-----RDELRRIKESLK-MERDQOFIATLREMIARD 2067
 Db 2709 -KEXAKELQE--ELLVKETNMTSLQKDLQSOVDHLAEAKLSILEKEDETEVQESKKAOM 2766
 Qy 2068 RQNHQVPEKELLS--DQOQHLMES-----LREKCS--RIKELLKRYSE 2107
 Db 2767 FEPLPIKLSKSIASCTDGTLLKISSNOTPQLLVKNAGIQINLQSECSSEVTEIISQFTE 2826
 Qy 2108 MDDHVECLNRLS-LDLE-KBIEFHRIMKKLVLYSYTKIKER-----QHECINKFEMD 2159
 Db 2827 KIEKQQLHAAIILDMESRHSISETETLKRHYVA--VOLLEKECGTLKAVIQCLRSKEGS 2884
 Qy 2160 FIDEVE-----KQKELL-----IKIQLHQQDCDVPS--RELRLDLKINMDLHI 2201
 Db 2885 SIPELAHSDAYQTEICSDSDSGSDWGQIYLTH--SQGFDIASBGRGSEESATDSFPKKI 2943

QY	2202	BEILKDPSEPPSIKTEFOVULS	-----NRKEMTOFLEWLNTR-----	FDIEKLK	2244
DB	2944	KGLLR	-----AVNEGQVLSLTSPSYDGEDHSIQOVSPFWLEERKAYINTISSUK	2995	
QY	2249	NGTOK-ENDRICOVNN	-----FNNRIIAIMNESTE-----	2278	
DB	2996	DLITKQOLQREAEVYOSSQSHESFSDWRGELLALAOQVFEERSVLLAARFELTALGTT	3055		
QY	2279	-----PEERSATTISKWEODLSLKEKNEK-LFNKYOTLTKTSILASGAQVNPPTQDNK	2329		
DB	3056	DAVGLLNCLFORIOEQGVEYQAAMECLQADRRSLLSEIOALHAQMGKRKITILKREQESE	3115		
QY	2330	NP	-----HVTSRAT-----OLTTEK--IRELNSLHEAK--	2356	
DB	3116	KPSQELLEYNIQOKSQOMLEQVLSMKDORATELOQSLSSEKMWAEKSELQAQKLEL	3175		
QY	2357	---ESAMHESKSLIKMOK-----ELEVNDIIAKLOAKVHESNKCLKTKETITQVLQ	2405		
DB	3176	ETTLKAQHKLKELEAFRLVKDKTDEVHLLNDTILASEQKSKSELOWALEKEKA	3229		
QY	2406	DKVALGAKPYKEETIEDLMKLVKIDLEKMKNAKEFEKEISATKATVYQKEVIRLLRENL	2465		
DB	3230	-KLRSEEROKEELEDLKFSL--ESQQRNLQ-----LNLLEPQQLLMSQOKI	3277		
QY	2466	RRSQOAOQTSVISHTPDPSNKPLTCGGSGIYQNTKALLKSEHRLKEKTSKLQOON	2525		
DB	3278	ESQRMLYDAQLSEE---QQRNLSEL-----QVLLSEKVRIREMSTLDR	3320		
QY	2526	E-----OLIKOKNELLNNQHLSNEVTKTWERTLK-----	2555		
DB	3321	ELHAQLQSGDGTGQSRPPLPSDDLKLKLOQLBKHSRIVELLNETEKYKLDLSLQTRQOM	3380		
QY	2556	---REAHQVTCENSPKSPKVYTGTSKKKOITPQCKERNLODPVPKESPKSCFFDRSK	2612		
DB	3381	EKDRQVHRKTL-----QTEQEAQTEGQKK-MHELOQSKVEDLQRLBEKRRQOVYKLDLEGQ	3434		
QY	2613	SLPSPHPVRYFDNSSLGLCPQVONAGAEVD-----SQPGPWHAASSCK	2655		
DB	3435	RLO-----GTMQBFQKQELREKREKRRRILYQNLNEPTWISDR	3476		

RESULT 11

```

US-10-171-311-6
; Sequence 6, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerish, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIORITY APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 3925
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-6

```

QY 816 -----DOEFONFKTLHMDPEQKYMVLEE-NERMNOBIV----- 848
 Db 1220 CEVNAEDKENS GDYI SERNEPELQDYRYEQDFQENMHTLLNKVTEYNKLLVLTQLSK 1279
 QY 849 -----NLSKEAQF-----DSSIGAL-----KTELSYK 871
 Db 1280 INGOOTDGMKLEGEENLPKEETEFLSIHSQMTNLEIDVNHKS KLSLQDLKTEKLEEQ 1339
 QY 872 TOELOKTRVQBRLEMEQLKEQLENRDSPLQTVBREKTLITELKQOTLEEVKTLTOEK 931
 Db 1340 VQELLESLSLQOQLKETE-----NYEAIHCLQKRLQAVSESTVPPSLPVDVSVITE 1393
 QY 932 DDILKQIQUESIQRDQLSDIHTVNNIDTQOLRNALBSLKHQOFTINTLSKI----- 987
 Db 1394 SDAORTMYPGSCVKKINDGTIFSGFGVKEETNIVKLE--KQYQOLBEEVAKVISM 1451
 QY 988 -----SEEVSRNLHMEENTGETKDEF-----QKWVGIDKKQOLEAKNTOTL--TADVK-- 1034
 Db 1452 STAFQOTELSRISGGKENTASSQAHAVCQOQHVFNEMKLSQDQIGQTFQTFETVDVKF 1511
 QY 1035 -----DNEIIEQQRKIFSLIOBKELQOQLESVIAEKEOL-----KTDLKENIEM 1079
 Db 1512 BEFKPLSKELGHEGKEILLNSDPHDIPESKOCVLTISEMFSKDTFFVRSIHDEISV 1571
 QY 1080 -TIENQOEELRLIGDELK-KQOEIVAKQKHAIKKEGEL-----SRTCDRLAEVEEKLEK 1132
 Db 1572 SSNDASRQMLNBEQLEDMRQELVROQEH--QOATELLRQAHMRQWRQREDOEQLOEE 1629
 QY 1133 SOOLOKQOQLNVOEE--MSEMQK-KINEIENLKN-----ELKN----- 1169
 Db 1630 IKRLNQLAQRASIDNENIUSRERVLLEBALKQI-SLAGEKLCCELRSNSTQONGN 1689
 QY 1170 -----KELTLBHMETERL--ELAQKL--NENVEEVKSTIKERKVLKELQKSF-----ETE 1215
 Db 1690 ENQGEVEEQTFKEKELDRKPEDVPPILSNRYALQKANNRLKILLEVKTAAVEETI 1749
 QY 1216 RDHLRGYI-----RETEATGL--QTEELKIAHILKHEQETIDELRSVSEKTAQIINTQ 1269
 Db 1750 GRHVILGILDRSSKSSASLIWRSEAAEASVSCVHEHTRVTDE--SIPSYSGSDMPEN 1806
 QY 1270 DLE-----KSHTKLOEEI-----PVLHEEQELLPNVKVSETOETMNE-LELLTEQ 1314
 Db 1807 DINMKSVTTEETELSORLVRSFGFAGTEIDPNEELMLNIS--SRLQAAVEKLELAISET 1864
 QY 1315 STTKDSTTLARIEMLRLINERFQBSQBEIKSLTKERDNLKIKEALEVKHQLKEHIRE 1374
 Db 1865 SSQLEHAKVTQELMR-----ESFRQKQATESL-----KCOEELRE 1901
 QY 1375 TLAKIQESQSQOQSLNKKKNETTKIVSEM-----EQFKPKDSALLRIEIMLGLSKRL 1430
 Db 1902 RLH--EESRAREQLAVELSKAEGVIDGYADEKTLPERQIQEKTDIIDRLQELLCASNRL 1959
 QY 1431 QESHDEMSVAKKDDLORLOEVLQSES-----DOLKENIKEIVAKHLET----- 1475
 Db 1960 QELEAEQOQIQEERELLGRKEMAKAEPVQQLQETKIMKEKLEJVVQCAEKVRDDL 2019
 QY 1476 BEELKVAHCLLKEQBETINELRVNLSEKETEISTIQOLEAINDKLONK----- 1524
 Db 2020 QKQVKALEIDVEEQVSRFIELE--QEKNTLMOLRQOQALEKQLEKRRFLDEQAIDR 2076
 QY 1525 -----IQEIVEKEQLNI--KQISEVENNVNELKQFKHRAKQDALSQIESKMLE 1573
 Db 2077 EHERDVFOQEIQKLEQQLKVVRFPQISEHQ--TREVEQLANHLKEKTDKCSSELLSKQE 2134
 QY 1574 LTNRLOESQEBI-----QIMIKEKEMKRVQF-----ALQIE 1605
 Db 2135 LQRIQERNESIEKLEFPVRLEQALLVSADTFQKVEDRKHFGAVEAKPELSLEVLQAE 2194
 QY 1606 RDQLKENTKIVAKMBSQEKQYQFLKMTAVNETQKMCBIEHLKEQFETOK-----LNLE 1661
 Db 2195 RDAIDRKEKIT--NLEEQLEQFRE-----ELENKNEVQQLHMQLEIQKKESTRLQ 2245
 QY 1662 NIETENIRLTOLHENLEEMRSVTKRDDLRSVEETL-----KVERDQKLENL 1709

Db 2246 ELBOEN-----KLFDDMEKLGIAKESDAMSTQDQHVLFQKFAQIIQKEVEIDQLNEQV 2301
 QY 1710 -----RETITRDLKQEBELKIVMHML--KEHQETIDKLRGIVSEKTN 1749
 Db 2302 TKLQOQLKITTDNKVIEEKNEILRDLETOIECLMSDOECVKRNRREEIEQLNEVIEKLOQ 2361
 QY 1750 EISNM-QK-----DLEHNDALKAQDLKIQEELRIAH----- 1780
 Db 2362 ELANIQKTSMNAHSISEEADSLKHQDQVVAEKLAEQOVETANEEMTFMKNVLKETNF 2421
 QY 1781 -----MHLKEQOETIDKLRGIVSEKTD-KLSNMQKDLNSNAKLEKIOELKANEH 1830
 Db 2422 KMQLIQELFSLAKRESVEKIIPIENSNNVAIDHLSKDKPELVILTE--DALKSLEN 2479
 QY 1831 Q-----LITLKKDVNETOKKVS-----EMEQLKKQIKDOOS----- 1860
 Db 2480 QTVFKSFEENGKSGIINLETRLLQLESTVSADLELTQCYKQIKDMQEQOQOQFETEMLOKK 2539
 QY 1861 -LTLKLEIENLMAOELHNLNLEEMKSVK-----ERDNLRRVEETLKLERDQ 1907
 Db 2540 IVNLQKIVEEKVAAALVSOILEAVQYAKFCODNOTISSEPRTNIQNLQ--LREDE 2596
 QY 1908 LKESLOETKARDLEIQO-----ELKTAEMLSKEHETVDKLRKISEKTIQISDIQKOLDKS 1964
 Db 2597 LGSDISALTRLRISELESQVVEHMTSLILEKEQVETIAEK--NVLKEKELKLELQKLEGN 2653
 QY 1965 KDELQKKIOELQKELQLLRVKEDVNMHKK--INEMQLKKQEPNVLCKCEMDFOL 2021
 Db 2654 EKKQREKERRSPQDVEVLKTTTELPHSNEESGFFNELALRAE---SVATKAEIASY-- 2708
 QY 2022 TKKLHLESLEEIRIVAKE-----RUELRRIKESLK-MERDQFIATLRMIARD 2067
 Db 2709 -KEKAELQOE-ELLVKETNMTSLQDLSQVRDLAEAKEKLSILEKEDETEVEQESKACM 2766
 QY 2068 RQNHQVKEPKRLLS--DGOQHLMES-----LREKCS--RIKELLKRYSE 2107
 Db 2767 FEPFLIKLSKSTASQTDGTLKISSNQTPQIILVKNAGIQINLOSCSESEVTEIISQFTE 2826
 QY 2108 MDHYECINRLIS-LDLE-KEIFHFRIMKKLYLSVYTKIEE-----QHECINKFEMD 2159
 Db 2827 KIEKQELHAAEILDMSRHSISETLKEHVVA--VQLLKEECGTLKAVIOCLSKESG 2884
 QY 2160 FIDEVE-----KQKELL-----IKTOHQDQCDVPS--RELDRDLKNQMDLHI 2201
 Db 2885 SIPELAHSDAYQTRIEICSSDSGSDWGQGIYLTH--SQGFDIASGREGSEESATDSFPKKI 2943
 QY 2202 BEILKDFESEFPFSIKTEFQOVLIS-----NPKEMTOFLEEWLNTF--FDIEKIK 2248
 Db 2944 KGLLR-----AVHNEGMOVLUTESPYSQDGHSTIQVSEPPLEERKAYINTISSIK 2995
 QY 2249 NGIQK-ENDRICQVNN-----FFNNRIIAIMNESTE----- 2278
 Db 2996 DLITKMQLOREAEVVDSSQSHESFSDWREGILLALQVFLERSVLAAFTTELTAQT 3055
 QY 2279 -----FERSATISKWEQDLKUKERNEK-LFKNYOTLKTSLASGAQVNPPTQDNK 2329
 Db 3056 DAVGLLNCLEQIOEQGVYQAAECLQKADRSILLSIQALHAQMGCRKITLKEQESE 3115
 QY 2330 NP-----HVTSRAT-----OLITTEK--IRELENSLHEAK-- 2356
 Db 3116 KPSOELLEYNIOQKOSQOMLEMQVELSNKORATELOEQLESSEKMMVVAELKSELAQTKLEL 3175
 QY 2357 ---ESAMHKEKSIIMKQK-----ELEVNDIIAKLQAKVHESNKCLEKTKETIQIVLQ 2405
 Db 3176 ETTLKAQHKLKELEAFLEVKDKTDEVHLLNDTLASEQKRSRELQWALEKEKA----- 3239
 QY 2406 DKVALGAKPYKEETIDLMKLVKIDLEKMKNAKEFEKISATKATVEYQKEVIRLLREN 2465
 Db 3230 -KLRSEERDEKEELEDLKFSL--ESQKQNLQ-----LNLLEEQKQLNESQOKI 3277
 QY 2466 RRSQQAODTSVISEHTDPOPSNKPILTCGGSGSIVONTKALILKSEHRLRKEISKLKQON 2525

Db 3278 ESQRMLYDAQLSEE---QGRNLEL-----QVLLSEKVRIREMSSTLDRER 3320
Qy 2526 E-----QLIKOKNELLNNQHLSNEVKTWKERTLK----- 2555
Db 3321 ELHAQLQSSDGTQOSRPPLPSEDLLKELQOLBEKHSRIVELLNETEKYKLDLSQTRQOM 3380
Qy 2556 ---REAHKQVTCENSPKSPKVTGTASKKKQITPSQCKERNLQDPVPKSPKSCFFPDSRSK 2612
Db 3381 EKDRQVHRKTL-----QTEGEANTGQKK-MHELOSKVEDLQRLBEKRRQVVKLDLGGQ 3434
Qy 2613 SLSPSPHPVRYFDSNLSGLCEPQVAGAESVD-----SQCPFWAASGK 2655
Db 3435 RLQ-----GIMQEFQKQELEREKESRRILYQNLNEPTTSLTSDR 3476

RESULT 12
US-10-370-685-100
; Sequence 100, Application US/10370685
; Publication No. US20030215903A1
; GENERAL INFORMATION:
; APPLICANT: Hyman, Paul
; TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional Elements
; FILE REFERENCE: NANF.P-004
; CURRENT APPLICATION NUMBER: US/10/370,685
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: 10/080,608
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 100
; LENGTH: 3911
; TYPE: PRT
; ORGANISM: human
US-10-370-685-100

Query Match 7.9%; Score 1055; DB 15; Length 3911;
Best Local Similarity 19.5%; Pred. No. 5.4e-44;
Matches 665; Conservative 616; Mismatches 1113; Indels 1016; Gaps 145;

Qy 26 TAQVYKTDNNVIYQVQDGSKFNFDRVPHGN-----ETTKVYERHATAPIIDSATQGYNGT 81
Db 307 TMOISFLQEKIKVYEMEQDK-----KVNSNKEEIOEKETIEELNTXIIIE----- 353
Qy 82 IFAYQOTASGKTYTMWG-----SEDLGVIPRAIHDFOKIKKFPDPREFLLRVSVMEIYN- 136
Db 354 -----EKKTLKLDKLTATADKLGLGELQEOIVQKNQEIKNMK-----LELINS 395
Qy 137 -----ETITDLICGTQKMKPLIREDVNRNVYVADLTBEVVYVTSWALKWITKGSKS 188
Db 396 KQKQESSEIEIKQLMGTVVEELQ-----KRNHKDSQFETDIVQRMQEQTORKLEQL-RAELD 450
Qy 189 RHYGE--TMMQORSRSHITFMILLESREKGEPSNC-----EGSVKYSHLNLDL 236
Db 451 EMYGOQIVQMKQELIRHQAQMEEMKTRHKGEMENALRSYSNITNEDQIKLMNVAINEL 510
Qy 237 -----AGSERAQGTACAGVRILKEGCNINRSLFILGOVIKLSDG-----QVGGFTINRD 285
Db 511 NTKLODTSNQKELKELGLILBEKALQRO---LEDIVELSFSEIQIQRARQIATAEQ 567
Qy 286 SKLTRILQNSLGNPKTRICTITP-----VSFDETLTALQFASATKYMNTPYVNEVST 340
Db 568 SKL-----NEAHSLSLTVEDLKBAIVSASERKELE-----LKHAEEVNYKI 610
Qy 341 DEALLKRYKEIMDLKQLEEVSL-----TFAQAMEKDQALAEKDLKQVQNEKIENL 397
Db 611 KLEMLEKEKNVLDRAESQAELERLTQLFSHEEELSCLKEDLEIHRINIEK----- 666
Qy 398 TRMLVTSSSLTLOELKAKRKRRTVWCLGKINKKNSYADQFNIPNTITTKTHK--LSI 455
Db 667 -----LKNLGIHYKQOID---GLQNEWSQKIETMQPE-KDNLITKQNQLILEI 711
Qy 456 NLLREIDES--VCSESDFVSNLDTL--SEIENWPNATKLLNQENITESELNSLRADYDNLVDL 513

Db 712 SKLKDLQQLSVNSKSEMTLQINELQKEIE-----ILRQE--EKEKGTLEQVEQL----- 760
Qy 514 YFQLRTEKEEMELKLEKNDLDE-----KQEMQLIHEISNLKLV-----KHREVVYNQOLE 574
Db 761 --QLKTLELQKQWKE--NDLQEKFAQLEAENSILKDEKKTLEDMLKIHTPVSQEERLIF 817
Qy 537 FEALERKTK-----KQEMQLIHEISNLKLV-----KHREVVYNQOLE 574
Db 818 LDSIKSKSDSVWEKIBILIEENEDLKQOCIOQNEIEKQRTVFAEKNFVYQELQ 877
Qy 575 NELSSSVKELLREKEDQIKKLOEYIDSKLENIKWIDLSYLESIEDPK-QMKQTLFD----- 629
Db 878 EETACLLKVKDDLEDS-KNQOELEYKSKLKAANEEL--HLQRINPTTVKMKSSVDFDRDKT 934
Qy 630 --AETVAL-----DAKRESAFLRSENLELEKEMKELATYTKQWENIDILYQSOLEA 678
Db 935 FVAETLEMGVEVVEKTTTELMEKLEVTREKLELSQRLSDLSQKQKHGISFTNEEBVKS 994
Qy 679 KK--KMQVDLE-KELQSAFNE-----ITKLTSLIDGKVPKDLLCNLELEKKITDLQ 726
Db 995 LKQEKQVSLRCELEIIINHNRANVQSCDTQVSSLDGVV---TMTSRGAESVSKVN 1051
Qy 727 KELNKE---VEE-----NEALREEVILLSELKSL-----PSEVRLRKBI 763
Db 1052 KSPGESKIMVEDKVPENMTVGEESKQEOQLILDHLPSTVKESSLRATQPSENDKLOKEL 1111
Qy 764 QD-KSEE-----LHI-----ITSEKDK---LFSVVH-KESRVQ 792
Db 1112 NVLKSEQDLRLQWEAQRICLSLVYTHVDQVREYMEKDKALCSLEELIFAQEEKIK 1171
Qy 793 GL-----LE-----EIGTKDLDATTOSNYKST----- 815
Db 1172 ELQKIHQLELQTKMTQETGDEKPLHLLIGLKQAVSECSYFQTLCSVLGEYTPALK 1231
Qy 816 -----DOEFQNFKTLHMDFFQKYRMVLEE-NERNNQIIV----- 848
Db 1232 CEVNAEDKENSQDIYSENEDELODYRYEVQDFQENMHMTLANKVTEYENKLLVLQTRLSK 1291
Qy 849 -----NLSKEAQKP-----DSSLGAL-----KTELSTYK 871
Db 1292 IWGQOTDGMKLEFGENLPKEETBFLSIHSQMTNLEDIDVNHKSKLSLQDLKTLKBEQ 1351
Qy 872 TOELOEKTRFVOERLNEQOLKEQLENRDSPLQTVREKTLITLQOOLEBEVTKLTQEK 931
Db 1352 VQELSLSSLOQQLKETEQ-----NYEAIHCLQKELQAVSESTVPSLPVDSVWITE 1405
Qy 932 DDLKQLOBSLQTERDQKSDIHDVTYNNMIDTQEOQLRNALESIKHQEFTINTLUSKI----- 987
Db 1406 SDAQRTMYPGSCVKKNIDGTIEFSGEFGVKEETNIVKLE--KQYQEOLEEEVAVIVSM 1463
Qy 988 -----SBEVSRNLHMBENTGETKDEF-----QOKMVGIDKKQDLEAKNTQTL-TADVK-- 1034
Db 1464 SIATAQOTSLRSISGKENTASSKQAHAVCOQEOHYFNEMKLSQDQIQIGFTETVDVVKPK 1523
Qy 1035 -----DNEITBQQRKIFSLIQEKNELOQMLESVIAEKEQL-----KTDLKENIBM 1079
Db 1524 EEFKPLSKELGHBGHEILLNSDPHDIPESKDCVLTISEMPSKDKTFIVRQSIHDEISV 1583
Qy 1080 -TIENQOEELRLGLDELK-KQBEIVAQEKNHAIKKEGEL-----SRTCDRLAEVEEKLEK 1132
Db 1584 SSMDASRQMLMNEEQLEDMRQELVRYQYQEH--QOATELLRQAHMRQEMERQEDQEOLEB 1641
Qy 1133 SOLOEKQOOLLNVQEE--MSEMQK-KINEIENLKN-----ELKN----- 1169
Db 1642 IKRLARQLAQRSSIDNENLVSRERVLLEELBALQKLSAGREKLCCELNRNSSTQNGN 1701
Qy 1170 -----KELTSHMEPERL--ELAQKL--NEWYEVKSTTKERKVLKEKQSF-----ETE 1215
Db 1702 ENQGEVEEQTPEKELDRKPEDVPPILSNERYALQKANNRLKILLEVVKTTAAVEETI 1761
Qy 1216 RDHLRGYI-----REIBATGL--QTKBELKIAHLKEHQETIDELRRSVSEKTAQINTQ 1269

Db 1762 GRVILGILDRSSKQSGASLIWRSEAEVSKSCVHEHRVRTDE---SIPSYSGSDMPRN 1818
Qy 1270 DLE-----KSHTKLOBEI-----PVLHEOEQLPNVKKVSETORTWNE--LELLTEQ 1314
Db 1819 DINMWSKVTEEGTSLQRLVRSFAGTETIPENELMNIS--SRQAQAEKLEJEAISSET 1876
Qy 1315 STYKDSITLARIEMERLURINEKFOESQETIKSLTKERDNLTKIEALEVHDQULKEHIRE 1374
Db 1877 SSQLEHAKVTQTELMR---ESFRQKQEAESL-----KQOELE 1913
Qy 1375 TLAKIOESQSKOBKSNMKEKNETTKIVSEM-----EOPFKDSALLRIEIMGLSKRL 1430
Db 1914 RLH--EESRAREQAVLSKAEGVIDGYADEKTLFERQIOETDIIIDRLEQELLCASNRL 1971
Qy 1431 QESHDEKSVAKEDDILQRLQEVLSSES-----DQLEKNIKEIVAKHLET----- 1475
Db 1972 QLEAEQOQOIEERELLRSRQKAEKAGVQEQQLLOETEKLMKEKLEVOQCAEKVRDDL 2031
Qy 1476 EELKVAHCLKQOEBETINELRVNLSEKTEISTIQKLEAINDKLQNK----- 1524
Db 2032 QOVKALEIDVEQVSRFIELE---QEKNTLMDLRQONQALEKQLEKMKRFLDEQAIDR 2088
Qy 1525 -----IQRIYKEBQLN1---KOLSEVOENVNELKQFKEHRKAKDSALQIESKMLE 1573
Db 2089 EHERDVFOEIQKLEOOLKVVPFPQISEHQ---TRVEQLANHLKBTDKCSELLLSKEQ 2146
Qy 1574 LTNRLQESQEBE1-----QIMKEKEEMKRVQE-----ALQIE 1605
Db 2147 LORDIQERNEETKLEFRVRELEQALLVSADTPQKVEDRGHFGAVEAKPELSLEVQLOAE 2206
Qy 1606 RQOLKENTKEIVAKMKESEKEQVQFKMTAVNETQBMKEIEHLKEQFTQK-----LNLE 1661
Db 2207 RQAIKREKEIT-----NLEQEQLEQPRE-----ELENKNEEVOQLHMOLEIQKKESTTRLO 2257
Qy 1662 NIETENIRLTQILHENLEMRSVTKERDDILRSVEETL-----KVERDQKLENL 1709
Db 2258 ELEQEN---KLFKODMEKLGAKIESDAMSTQDQHVLFKFAQIIQOEKEVEIDQLEQ 2313
Qy 1710 -----RETITRDLKQELKIVHMH1--KEHQBTIDKLKGIYSEKTN 1749
Db 2314 TKLOOLKITTONKVIIEKNELIRDLTQIECLMSQOECKVRNREIEIQLENEVEKLOQ 2373
Qy 1750 EISNM-QK-----DLEHSDALKAQDLKIQEELRIAH----- 1780
Db 2374 ELANIGQKTMNAHLSSEADSLKHQLDVVIKALEQOQVETANEEMTFMKNVLKETWF 2433
Qy 1781 -----MHLKEQOETIDKLKRGIVSEKTD--KLSNMQKDLNSNAKLOEKLOELKANEH 1830
Db 2434 KMNQLTQELFKLRESEVKEIQSIPENSUNVAIDHLSKDKPELEVVLITE--DALKSLEN 2491
Qy 1831 Q-----LITLKQOVNETQKVS-----EMEQKKQIKDQS----- 1860
Db 2492 QTYFKSFENGKSGIINLETRLLQLESTVSAKDLELTQCYKQIKDMQEQGQOPEMLOKQ 2551
Qy 1861 -LTLKSLIENLNAQELHENLEEMKSVK-----ERNLARVEETLKLERDQ 1907
Db 2552 IVNLOKIVEEKVAAALVSOIQLEAVQYAKFCQDNQTISSEPERTNIQNLQ--LREDE 2608
Qy 1908 LKESLQETKARDLEIQ---ELKTARMLSKKHETVDKLRKISEKTIQISDIQKLDKS 1964
Db 2609 LGSDISALTALISELSQVEMVHTSLILEKEQVEIAEK---NVLEKKEKLELLEKLEGN 2665
Qy 1965 KDELOKQIQELQKLEQLLRVKEDVNMGHK---INEMEQLKKQPEPNVLCXCMDFOL 2021
Db 2666 EKKQREKKGKSPQDVEVLKTTTELPHSNEESGFPNEALRAE---SVAKELASY-- 2720
Qy 2022 TKLHESLEERIVAKE-----RDELRIKESLK-MERDOFIATLREMIARD 2067
Db 2721 -KEKAEKLOE-ELLVKETNMVTSLOKDLQSVRDHLAAEAKELSILEKEDETEVEQSKKACM 2778
Qy 2068 RQNHQVPEKRLSDGQOQHLMESLREKCSRILKELLKRYSEMDHDHVECLNRLSLDEKIE 2127
Db 2779 FEPLPIKLSKIASQTDGTLKISSNQTPQI--LVKNAG-----IQINLOSECS 2825

Qy 2128 FHRIMKKLVSVTVTKIKEBQEHCKNFEMDFIDEVEK-QKELLIKIQLHQDQCVPSR 2186
Db 2826 SEEVETIISOFTEKIERMQELHAAEILDMESRHSISETETLKRHHVAVQLLKEEGTLKA 2885
Qy 2187 ELRDLKL-----NQMDLHIEEILKDPSES---EFP 2214
Db 2886 VIQCLRSKVFVFNMCFTSLCDSGSDWGQGIIVLTHSQGFDI-ASEGRGESESATDSFP 2944
Qy 2215 -----SIKTEPQOVLVS-----NRKEMTQFLEEWLNT----FDEIKLKNGIQK 2253
Db 2945 KTIKGLLRVAHNEGQVLSLTPESYPDGEDHSIQOVSFPEWLEKAKYINISSLKDLTK 3004
Qy 2254 -ENDRICQVNN-----FFNRIIAIMNESTE----- 2278
Db 3005 MQLQREAEVYDSSQSHESFSDWRGELLALQOVFLERSVLLAAFRTELTAFTGTTDAVGL 3064
Qy 2279 ---FEERSATISKWEODLKSLEKNEK-LFKNYQTLKTSLSAGAQNVPNTQKNP--- 2331
Db 3065 LNCLEQRIQEOGVEYQAMECLOKADRRSLSEIQALHAQMNGRKITLKREQSEKPSQE 3124
Qy 2332 -----HVTSRAT-----OLTTEK---IRELENSLHEAK-----ES 2358
Db 3125 LLEYNIQKQSQWLEMQVELSSMKDRATELOEQLSSEKVMVAELKSELATQKLEETTLK 3184
Qy 2359 AMHESKIIIMQK-----ELEVNDIIAKLOAKVHESNCKLEKTKETIQVLODKVAL 2410
Db 3185 AQHGHKELEAFRLVCKDTDEVHLINDTLASEQKKSRELQWALEKEKA-----KLGR 3237
Qy 2411 GAKPYKEETEDUKMLVKIDLEKMKNAKEFEKEISATKATVEQKEVIRLLRENLRSSQ 2470
Db 3238 SEERDKEELEDLKFSL---ESQKQRMQ-----LNLLEQOQKOLLNESQOKTESQRM 3286
Qy 2471 AOTSVISHTDPPSNKPLTCGGSGIVQNTKALIKSEHIRLEKEIKSLKQONE--- 2526
Db 3287 LYDAQILSEE---QGRNLEL-----QVLESEKVRIREMSSTLDRERLHAQ 3329
Qy 2527 -----QLIKQKELLNNQHLNNEVKTWERTLK-----RE 2557
Db 3330 LOSSDGTQSRPLPSBELLKELQLEKHSRIVELLNETEKYKILDSLTQRMQEKDRQ 3389
Qy 2558 AHQVTCENSPKSPKVTGTASKKKOITPSQCKERNLODPVPKESPKSCFFDSRSKSLPSP 2617
Db 3390 VHRKTL-----QTEQEAETEGQKK-MHELOSKVEDLOQLEEKQOQVYKLDLEGORLQ-- 3441
Qy 2618 HPVRYFDNSSLGLCEPQVQNAEBSYD-----SQGPFMHASSGK 2655
Db 3442 -----GIMQEPQKQELEREKESRRILYQNLNEPTTWLSITSDR 3480

RESULT 13

US-10-408-765A-1839
; Sequence 1839, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Watcock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1839
; LENGTH: 3911
; TYPE: PRT
; ORGANISM: Homo sapiens

Fri Aug 6 10:49:04 2004

US-10-408-765A-1839

Query Match 7.9%; Score 1055; DB 16; Length 3911;
 Best Local Similarity 19.5%; Pred. No. 5.4e-44; Indels 1016; Gaps 145;
 Matches 665; Conservative 616; Mismatches 1113;

QY 26 TAQVYKTDNNVYQVDSKSFNFDRVFGN-----ETTNNVYEEIAAPIIDSIAIOYNGT 81
 DB 307 TMOISFLQEKIKYEMEQDK-----KVNSNKEIEQEKETIIEELNTKIIIE----- 353
 QY 82 IFAYGOTASGKTYTMMG-----SEDLGVIPRAIHDFQIKTKPPDRFLLRVYMEIYN- 136
 DB 354 -----EKKTLKDKLTADKLLGELQEQIVQKNOBKNNK-----LELTNS 395
 QY 137 -----EYITDLCTQKMKPLIREDVNRNVYVADLTVEVYVTSEMALKWITKGEKS 188
 DB 396 KQKERSSEBIEKOLMGTVBELQ-----KRHKDSQFETDIVORMEQETQRLQEI-RAELD 450
 QY 189 RHYGE--TKMNRSSRSHTIPRMILEGREKEGPSNC-----EGSVKVSHLNLVDL 236
 DB 451 EMYGQOIVQMKQELIROMAQMEEMKTRHKEMENALRSYSNITVNEDOIKLNNVAINEL 510
 QY 237 -----AGSRAAQTGAAGVRLKGCNINRSLFQIVIKLSDG-----QVGGFINYRD 285
 DB 511 NIKLQDTSQEKLEKEELGLILEKCALQO---LEDLVEELSFRBQIQARQTIABQE 567
 QY 286 SKLTRILONSLGGNPKTRIICITIP-----VSFDETLTALQFASTAKYMKNTPVVNEVST 340
 DB 568 SKL-----NEAHKSLSTVBDLKABIVSASESRKELE-----LKHEAEVTNYKI 610
 QY 341 DEALLKRYKIBMDKKQLEBVSLE--TRAQAMEKDQALQLEEKDOLLOKQVQNEKIENL 397
 DB 611 KLEMLEKKNVLDORMAESQAELERLTQLFSHEBELSKLEDLEIHRINIEK----- 666
 QY 398 TRMLVTSSSLTQOELKAKRRVTCUGINKNNKSNYAQFNIPNTIITKH--LSI 455
 DB 667 -----LKNLGHYKQOIQID---GLONEMSQKLETWQFE-KDNLLITKQNLILEI 711
 QY 456 NLLREIDES-VCSESDVSNTLDTI--SEIWNPAKLANQENIESELNRAADYNLVID 513
 DB 712 SKLDLOOSLVNSSEMTLOINELQKEIE-----ILRQE--EKEKGTQEQVQEL--- 760
 QY 514 YEQLRTEKEEMELKKEKNDLDE-----NDOKEFPAQLAENSILKDEKTKLEDMLKIHTPVSOERLIF 817
 DB 537 FEALERKTK-----KQDEMQLIHEISNLXNV-----KHREYVNOBLE 574
 QY 818 LDSIKSKSDSVWEKEIEIIBENEDLKQOCTQLNEEIEKORNTFSFAKNPEVNVQELQ 877
 QY 575 NELSSKVELLREKEDQIKKQVDSQKLENTKMDLSYSLIESIEDPK-QMKQTLFD--- 629
 DB 878 EYACLLKVKDLEDS-KNKQELVYSKULKALNEEL--HLQRIPTVVKMSVDFDEK 934
 QY 630 --AETVAL-----DAKRESAFILSENLEKEMKELATYQKQENDIOLYQSLEA 678
 DB 935 FVAETILEMGEVVEKOTTELMEKLEVTREKLELSQRLSDLSQKQKHGEISFLNEEVKS 994
 QY 679 KK-KMQVDLE-KELOSARNE-----ITKLTSLIDGKVPKOLLNLEEGKTTDLQ 726
 DB 995 LKQKEQVSLRCRELEIIINHNRAENVQCDTQVSSLLDGW---TWTSGRAGSVSKYN 1051
 QY 727 KEINKE---VRE-----NEALREEVILLSELKSL-----PSEVERLRKBI 763
 DB 1052 KSGFGEKIMVEDKVSFENMTVGEESKQROLILDHLPSTVKESSLRATQPSENDKLOKEL 1111
 QY 764 QD-KSBE-----LHI-----ITSEKDK-----LFEVNH-KESRVQ 792
 DB 1112 NVLKSQNDLRLQWEAQRICLSIVSYTHVDQVREYMEKDKALCSLKEELIPAEKBIK 1171
 QY 793 GL-----LE-----EIKTKDDLATTOGNYKST----- 815
 DB 1172 ELOKIHOLELQTKTQETGEGKPLHLILIKLOKAVSECSYFLQTLCSVLGEVYTPALX 1231

QY 816 -----DQBFQNFKTLHMDPEQYKMLVLE-NERMNOEIV----- 848
 DB 1232 CEVNAEDKNSGDIYSENEDPELQOYRYEVQDFQENHLLNNKYTEYNKLLVLQTLKSL 1291
 QY 849 -----NLSKEAQKF-----DSSLGAL-----KTLSYK 871
 DB 1292 INQOQTDGKMLPEGENLPKEETEFLSIHSQMTNLEDIDVNHKSLSLQDLKTEKLEEQ 1351
 QY 872 TQBLQEKTRVQERLNEFQLEQLENRDSPLQTVREKTLITTEKLOOTLEEVKTLTQEK 931
 DB 1352 VQELSELSLSLOQLKXETQ-----NYEAEIHLQKRLQAVSESTVPPSLPVDVSVITE 1405
 QY 932 DDUKQLOESQIIBRDQKSDIHTVNMNIDTQOLRNALLESKQHOETINTLSKI----- 987
 DB 1406 SDAQRTWYPGSCVKKNIDGTIEFSGFVGKEETNVKLE--KQYQOLEEVAKVISM 1463
 QY 988 -----SEVSRNLHMEENTGETKDEF-----QOKMVGIDKKODLEAKNTOTL-TADVK-- 1034
 DB 1464 SIAFAQOTELSRISGGKENTASSQAHAVCQEQHYFNEMKLSQDQIGFOTFTVDVKFK 1523
 QY 1035 -----DNEITROOKIFSLIOEKNELOQMLSVIAKEQOL-----KTDLENLEM 1079
 DB 1524 EEPKPLSKELGEGHKEILLSNDPHDIPESKDCVLITISEMFSKDFTFIVRQSIDHLSV 1583
 QY 1080 -TIENQEBELRLGDELK-KQOEIVAOEKHAIKKEGEL-----SRTCDRLAEVEEKLEK 1132
 DB 1584 SSMDASQMLNEEQLEDMQELVRQYQEH--QOATELLQAHNMQRQERQEQEQEE 1641
 QY 1133 SQLOEQKQOQLNVQEE--MSEMOK-KINIEINLKN-----ELKN----- 1169
 DB 1642 IKRLNRQLAQSSIDNENLVSESERVLELEALQKLSAGREKLCCELRNSSTQTONGN 1701
 QY 1170 -----KELTLEHMETERL--ELAQL--NENYEEVKSITKERVKLEKQSF-----ETE 1215
 DB 1702 ENQGEVEEOTPKEXELDRKPEDVPEILSNRYALQKANNLLKILLSEVNVKTTAAVESTI 1761
 QY 1216 RDHLRGYI---REIATGL--QKKEELKIAHILKHEOETIDELRRVSEKTAQIINTQ 1269
 DB 1762 GRHVILGILDRSKSQSSASLIWRSEAEASVKSCHVEHTRVTDE--SIPSYSGMDPRN 1818
 QY 1270 DLE-----KSHTKQIEBI-----PVLHEEQELLPNVKVSQETQETMAE-LELLTQ 1314
 DB 1819 DINMWSKVTEGTELSQLRVSFAGTAGTEIDPENELMLNIS--SRLOAAVEKLELAIST 1876
 QY 1315 STTKDSTTLARIEMERLRLNEKFQESQIEKSLTKERDNLKTIKEALEVHMDOLKEHIRE 1374
 DB 1877 SSQLEHAKVTQELMR---ESFRQKQEBATESL-----KCOEELRE 1913
 QY 1375 TLAKIQESQSQEQSLNMKKNEDNTEKIVSEM---EQFKPKDSALLRIEIMGLSKEL 1430
 DB 1914 RIJH--EESRAREQLAVELSKAGVIDGYADEKTLFERQIQEKTDIIDRLQELLCASNRL 1971
 QY 1431 QESHDEMSVAKEKDDQLQRLQEVLOSES-----DQKENIKEIVAKHLET----- 1475
 DB 1972 QELEAEQQOIQEERELLSSROKEMAKAEPVQQLQETELKMEKLEVQCOAKVRRDL 2031
 QY 1476 ERELKVAHCLKEQETINELRVNLSKETETSTIQKOLEANDKLNK----- 1524
 DB 2032 QKQVKALEIDVEEQVSRFIELE---QEKNTLEMDLRQQQALEKQLEKMKRKFDEQAIDR 2088
 QY 1525 -----IQEIVKEKEQLNI---KQISEVQENVNVELKQFKERKAKDSALOSIESKMLE 1573
 DB 2089 EHRDVFQEQIQLKLEQOLKVVPRFPQISEHQ--TREVEQLANHLKEKTDKCSSELLSKQ 2146
 QY 1574 LTNRLQESQEEI-----QIMKEKEEMKRVQE-----ALQIE 1605
 DB 2147 LQORDIQERNEEIEKLEFRVRELEQALLVSADTFQKVEDRKHFGAVAKPELSLEVLQAE 2206
 QY 1606 RDOLKENTKEIVAKMESQEKYQFLKMTAVNETQEKMKCEIEHLKEQFETQK---LNLE 1661
 DB 2207 ROAIDRKEKEIT---NLSEOLEQFRE-----ELENKNEEVQQLHMQLEIOKKESTRLQ 2257

QY	1662	N I E T E N I R U T Q I L H E N L E M R S V T K E R D L R S V E T L	1700
Db	2258	E L E O E N --- K L F K D M E K L G L A I K E S D A M S T Q D Q H V L F G K F A Q I O E K E V E I D Q L N E Q V	2313
QY	1710	--- R E V I T R D L E K O E B E L K I V H M L --- K E H Q E T I D K L R G I V S E K T N	1749
Db	2314	T K L Q O O L K I T T N K V I E K N E L I R D L E T Q I E C L M S D Q E C V K R N E E E I Q L N E V I E K L O Q	2373
QY	1750	E I S N M - Q K --- D L E H S N D A K A Q D L K I Q E E L R I A H ---	1780
Db	2374	E L A N I G O K T S M N A H S L S E E A D S L K H Q D V I A E K L A E Q Q V T A N E M T M K N V K E T N F	2433
QY	1781	--- M H L K E O E T I D K L R G I V S E K T D - K L S N M K O L E N S N A K L O E K I O E L K A N E H	1830
Db	2434	K M N Q L T O E L F S L K R E R E S V E K I Q I S P E N S V N V A I D H L S K D K P E L V V L T E - D A L K S L E N	2491
QY	1831	Q --- L I T L K K D V N E T O K K V S --- E M E Q L K K O I K D Q S ---	1860
Db	2492	Q T Y P K S F R E N G K S I I N E T L L Q L E S T V S A K O L E T Q C Y K I O M O E Q O G F E T E M L O K K	2551
QY	1861	- L T I S K L E I E N L N A Q E L H E N L E M K S Y M K --- E R D N L R R V E E T L K L E R D Q	1907
Db	2552	I V N L Q K I V E E K V A A L V S Q I Q L E A V Q E Y A K F C D N Q T I S S E P E R T N I Q N L N Q --- J R E B E	2608
QY	1908	L K E S I O E T K A R D I E I Q --- E L K T A R M L S K E H K E T V D K L E K T I S E K T I Q I S D I Q D K D L R S	1964
Db	2609	L G S D I S A L T L R I S E L S Q V V E M H T S I L E K E Q V E I A E K --- N V L E K E K L L E Q L K L E G N	2665
QY	1965	K D E L Q K I Q I O E L K O L L R K R E D V N M S H K --- I N E M E O L K K Q F P E N V I C K E M D N F Q L	2021
Db	2666	E K X Q R E K E K R S P O D V E V L K I T T E L F H S N E S G F F N E L A R A E --- S V A T K A E T A S Y ---	2720
QY	2022	T K L H E S L E E I R I V A K E --- R D E L R R I K E S L K - M E R D Q I F A T L R E M I A R D	2067
Db	2721	- K E A E K L Q E - E L I V K E T N M T S L Q K O L S Q V R D H L A E A K E K I L E K E D E T E V Q E S K A C M	2778
QY	2068	R O N H O V K P E K L I S D G O H L M E S I R E K C S R I K E L L K R Y S E M D H Y E L N R L S D L D E K E I T	2127
Db	2779	F E P U P I K L S I A S Q T D G T L K I S S N Q T P O I --- L V K N A G --- I Q I N L Q S E C S	2825
QY	2128	F H R I M K K L K V L S V Y T V I K E E Q H E C I N K F M D F I D E V E K - O K E L L I K I O H L Q O O C D V P S R	2186
Db	2826	S E E V T E I I S O F T E K I E K O Q E L H A E I L D M E S R H I S E T E T L K R E H Y V A V O L L K E E C G T L K A	2885
QY	2187	E L R D L K I --- N Q N D H I E B I L D F D S E S --- B F P	2214
Db	2886	V I O C L R S K E Y F G Y N M C F S T C D S G D W G O G I V L T H S Q G F D I - A S E G R E E S E S A T D S F P	2944
QY	2215	--- S I K T E F Q O V L S --- N R E M T O F L E W L N T R --- F D I E K L N G I O K	2253
Db	2945	K K I G L L R A V H N G M Q V L S T E S P Y S D G R H S I Q Q V S E P W L E R K A Y I N T I S S L K D L I T K	3004
QY	2254	- E N D R I C Q V N N --- F F N N R I I A I M N E S T E ---	2278
Db	3005	M Q L O R A E V Y D S Q S H E S F S D W R G E L L A L Q O V F L E E R S V L L A A P R T E L T A L G T T D A V G L	3064
QY	2279	--- P E R S A T I S K E W O D L K S K E R N E K - L F N Y O T L T S L A S G A Q V N P T T Q D N K N P ---	2331
Db	3065	I N C L E O R I O B Q V E Y O A A E C L Q A D R S L L S E I Q A L H A Q M G R K I T L K R Q E S E K P S Q E	3124
QY	2332	--- H V T S P A T --- O L T T E K --- T R E L E N S L H E A K --- E S	2358
Db	3125	L L E Y N I Q O K S O M L E M Q V E L S M K O R A T E L O Q L S E R K M V A E L S E L A O T K L E T T I K	3184
QY	2359	A M H E S K I I K M O K --- E L E V N D I I A K L O A K V H E S N K C L E K T K E T T I Q V L Q D K V A L	2410
Db	3185	A Q H K L K E L A F R L E V K D K T D E V H L L N D L A S E Q K S R E L O W A L E K E A --- K I G R	3237
QY	2411	G A K P Y E E E D L K M L V K I D L E K M K N A K E F E I S A T K A T V E Y Q K E V I R L L R E N L R S Q Q	2470
Db	3238	S E E R D K E E L D L K F S L --- E S O K O R N L Q --- L N L L L E O Q K L L N E S Q K I E S O R M	3286
QY	2471	A O D T S V I S H T D P O P S N K P L T C G G S G I V O N T K A L I L K S E H I R L E K I S K U K O N E ---	2526

```

Db      3287 LYDAQISE-----QQRNLEL          :|::||:::||::||::||::||::||::||::||
QY      2527 -----OLIKOKNELSNNHLSNEVKTWKERTLK-:|::||::||::||::||::||::||::||::||
Db      3330 LQSSDGTGSRPLPSPEDLLKELQKLEEKHSRIVELLNETEYKYKLDSLOTROQMEKDQ 3389
QY      2558 AHKQVTCENSPKSPKVTTASKKIQTSPQCKERNLODPVPKSPKSCFFDRSKSLPSP 2617
Db      3390 VHRKTL-----QTEQEAANTEQQKK-MHELOS KVEDLQRLQEERQOYYVKLDLEGORLQ-- 3441
QY      2618 HPVRFDNSSGLCPVEQNAGAESVD-----SOPGPWHASSGK 2655
Db      3442 -----GIMQEFQKQELEREKRESRRILYQNLNEPTTWSLTSDR 3480

RESULT 14
US-10-080-608A-11
; Sequence 11, Application US/10080608A
; Publication No. US20030198956A1
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/10/080,608A
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-080-608A-11

Query Match              7.9%; Score 1052.5; DB 14; Length 3878;
Best Local Similarity    19.5%; Pred. No. 7.1e-44;
Matches 669; Conservative 603; Mismatches 1069; Indels 1089; Gaps 144;

QY      26 TAQVYKTDNNVIYQVDGSKSFNFDFVHGHN---ETTANNVVEETAAPTIDSAIOGYNGT 81
Db      307 TMQISFLOEKIKYVEMQDK-----KVENSKEEQETIIIEELNTKIIIE----- 353
QY      82 IFAYGOTAGKTYTMNG-----SDHLGVTPRAIHIFQIKIKFPDPREFLLRVSYMEIYN- 136
Db      354 -----BKTLELKDULTADKLGELQEQIVQKNQEIKNWK-----LELTNS 395
QY      137 -----ETITDLCTGMKPLIREDVN RVVADLTEEVVVTSEMAKWITKGSKS 188
Db      396 KQKEROSEEIKQLMGTVELQ-----KRHKDSQPETDIVQRMEOETQTKLEQL-RAELD 450
QY      189 RHYGE-TKMGNRSSRSHTIFRMILLESREKGPSNC-----EGSVKSHNLVDL 236
Db      451 EMYGQIQVMKQELIRQHVAQMEMKTRHKGEMALRSYSNITVNEDOIKLMMVAINEL 510
QY      237 -----AGSERAAQTGAAGVRLKEGCINNRSLFILGOVKKLSDG-----OVGGFINYRD 285
Db      511 NIKLQDTNSQKEKKEELGLLILEKCALQRQ----LEDIVELSRSREQIQARQTIAEQ 567
QY      286 SKLTRILQNSLGNPKTRIICHTIP-----VSFDETLTALOPASTAKYMKNTPYVNEVST 340
Db      568 SKL-----NEAHKSLSLTVEDLKAEIVSASESRKELE-----LKHEAEVTNYKI 610
QY      341 DEALLKRYBEITMDLKQLEEVSLE---TRAQAMEKQOLAQLLEKOLLQKVONEKIENL 397
Db      611 KLEMLEKENAVLD RMWSSQAEBULRUTQLFSHEBELSKLKDLEIETRINTEK----- 666
QY      398 TRMLVTSSSLTIQBELKAKRRRVVTCWGKLNKMKNSYADQFNIPNTITKTHK--LSI 455
Db      667 -----LKONLGIHYHQOID-----GLQNMESQKETWQFE-KONLITKQNLLEI 711
QY      456 NILREIDS-VCSDESDFNSTLDTL-GSEIWNPATKLIINOENIESELNSLRADYDNVLVD 513
```

Query Match 7.9%; Score 1052.5; DB 14; Length 3878;
Best Local Similarity 19.5%; Pred. No. 7.1e-44;
Matches 669; Conservative 603; Mismatches 1069; Indels 1089;

Qy	26	TAQVYWKTDNNVIVQVDGSKSFNEDRVFHN-----ETTKNVEEETAAPIIDSALQVNGT	81
Db	307	TWOISFLQEKIVYEMEQDK-----KVNSNKEEIQEKETIEELNKLIEE-----	353
Qy	82	IFAYGQTASGTYTWMG-----SEDLGWIPRAIHDIPOKIKKPPDEFLLRVSYMEIYN-	136
Db	354	-----EKKTELKDLKLTATDKLLGELQEQIVQNKQEIKNMK-----LELTNS	395
Qy	137	-----ETITDILLCGTQKMKPLIIRDVNRNVYADLTREVVTSEMALKNWTKGKKS	188
Db	396	KQKERQSEETIKQLMGTVEELQ-----KENHKDSQFETDIVORMEQOTKRLQL-RAELD	450
Qy	189	RHYGE--TKMNQRSRSHTIFRMILESKEKPEPNC-----EGSVKYSHLNVLDL	236
Db	451	EMYGOQIVOMQELIQRHQAOMEEMKTRHKEMENALRSYSNITVNEDOIKNVAINEL	510
Qy	237	-----AGSERAAQTGAAGVRLKEGCNINRSLFILGOVKKLSDG-----QVGGFINVD	285
Db	511	NIKLDQTNQSKEKLEELGLILEEKALQRO-----LEDLVEELSPSREQIORAQOTIAEO	567
Qy	286	SKLTRIILQNSLGGNPKTRIICTITP-----VSFDETTALQFASAKYMKNTPTPYNEVST	340
Db	568	SKL-----NEAHKSLSTVEDLKAETIVSASESRKELE-----LKHEAEVTNYKI	610
Qy	341	DEALLKRYKREIMDLKKOLEEVSLE-----TRAQAMEKQOLAQLLEKDLQKVONEKIENL	397
Db	611	KLEMLEKEKNAVJORMAESQAELELURTOQLFSHBEELSXLKEDLEIHRINTEK----	666
Qy	398	TRMLVTSSSLTLOQLKAKKRRVTRCGLGINKNMKSNYADQFNIPTWITTKHK--LSI	455
Db	667	-----LXONLGIHVKKQID-----GLQENMSQKIETWQFE-KONLIITKQJLILEI	711
Qy	456	NLUREIDES-VCSESDFVSNLTDLT-SEIEWNPATKILNOENIESELSNFRADYDNLVLD	513

Db	712	SKLQOLQOVLNKSSEMTLOINELQKEIE-----ILRQE-EKEKGTLEQVQEL---	760
QY	514	YEQLRTEKEBEMELKKEKNDLDE-----	536
Db	761	--OLKTLELLEKOMKEKE--NDLQEKFAQLEAENSILKDEKKTLEMDLKIHTPVSQERELLF	817
QY	537	FEALERKTK-----KQEMQLIHEISNLKLV-----KHREVVYNQDLE	574
Db	818	LUSIKSKSDSVWEKIEIILIEENEDLKQOCIQLNNEIEKQRTFSPAENFVNVQELQ	877
QY	575	NELSSKVELLREKEDQIKLQYIDSKLENIKMDLSYSLESTEDPK-QMKQLFD----	629
Db	878	EYVACLLKVKDDLEDS-KNKQELLYSKLKALNEEL--HLQRINPTTVKMSVFDDEK	934
QY	630	--AETVAL-----DAKREAFILSENLEKEMKELATTYKQOMENDIQLYQSOLEA	678
Db	935	FVAETLEMGVFEKDTTELMEKLEVTREKLELSORLSDLSBQKHGHEISFLNEEVS	994
QY	679	KK--KMOVDLE-KELQAFNE-----ITKITSIDGKVPKDLLCNLELEGKITDLO	726
Db	995	LKEKEQVSRCRELEIIINHNAENVQSCDTQVSSLLDGV--TWTSRGAEGSVSKVN	1051
QY	727	KELNKE-----VEE-----NEALREVILLSELKSL-----PSEVERLKEI	763
Db	1052	KSFGEESKIMVEDKVSFENMTVGBESQEQIILDLPLSVTKESSLRATQSPENDKLOK--	1109
QY	764	QDKSEELHIITSEKDKLFSEVVHKSRSVQGLLEIGTKDLDATTQSNYKSTQDEFQNF	823
Db	1110	-----ELNVLKSEOND-----RLQMAQRI-----CLSLVYST-----	1138
QY	824	TLHMDFEQYKWLNEERNMQEIVNLSKEAKFDSLSGALKTSLYKTOBLOKTRVQ	883
Db	1139	--HVDQVREYM-----ENEK-----DKALCSLKEELIPAA-----BEKIKELQ	1174
QY	884	ERINEMQLKEQLENRDSPLQTVREKT-----LITEKLOQTLLE-----VKTLTOE	930
Db	1175	KIHOLE-----LQTMKTQETGDEGKPLHLGLIGLOKAVSECECSYFIQTLCSV	1221
QY	931	KDDLKQLESLOIERDQKSDIHDVTVMNIDTQEQLRNALSLSKHQOETINTLSKISSE	990
Db	1222	LGYYTTPALKCEVNAED--KENSQDYISENED--PELQDYREYVQDFQENMHTLKNKVT	1278
QY	991	VSRNLHM-----BENTGETNDEFQQRMGVIGDKODLEAKNTOTLT	1030
Db	1279	YNKLLVLQTRLSKIMGQOTDGMKLEFGEENLPKBTETFLSHSQMNLDEDVNVHKSLS	1338
QY	1031	A--DVKDNEIITQOKKIPSLI-----QEKNEILOOMLES-----	1061
Db	1339	SLQOLEKTKLEFQVQELESLISSLOOQLKETEQNYEABIHCLQKQLQAVSESTVPPSLPV	1398
QY	1062	---VIAEKEBQKT---DLKENIEMTIE-----NOBEL-----	1087
Db	1399	DSWITESDAORTMYPGSCVRKNIDGTITFSGFEGVKBEETNIVKLEKQYQOLSEEVAK	1458
QY	1088	-----RLIGDE-----LKKQOBIVAQEKNAHAIKGBELSRUCDLRA----	1123
Db	1459	VIVMSIAFAQOTELSRISGGKENTASSKQAHAVQOEQHYF-----NEMKLSQDQIGFQTF	1515
QY	1124	EVEKLEKKSQOLO-----EKQOQLNVQOEBEMSEMQKKNIE	1161
Db	1516	ETVDVVKFEKPLSKELGHEGKEILLNSDSDPHDIPESKDCVLITISEEMFSDKDTFIVRQ	1575
QY	1162	NLKNELK-----NKELTLEHETERLELAQKLNENYEVKSTIKERKVLKELQKSFETE	1215
Db	1576	SHIDEISVSMOASRQMLNNEQLE--DMQELVRQYQEHQATQRSSIDNE--NLVSE	1630
QY	1216	RDLRGYIREIATGLQTKKEIKIAHIHKEHQETIDELRRSV-----SEKTAQIINTQ	1269
Db	1631	RE--RVILLELEA-----LKQLSLAGREKLCCLERNSTQTONGENQGVEEQ	1678
QY	1270	DLEKSHTKLOEETP--VLHEEQE-----LLPNVKYSETQETMNE--LELLTE	1313
Db	1679	FXEKELDRKPEDVPPPEILSNERYALQKANNRLIKILLEVVVKTAAVEETIGRHVLGILDR	1738
QY	1314	QSTTKDSTTL-----ARI-----EMERLRN-----EKF	1337
Db	1739	SSKSSSASLIRSEAEASVKSHHEHTRVTDSTPSSVSGSDMPNDINMWSKVTEEGT	1798
QY	1338	QESQEBIKS-----LTKERDNLK-TIKEALEVKHDQKHEHRETIAKIQESOSKQ----	1386
Db	1799	EJSORLVRSGFAGTEIDPENEELMLNISRLOAAVEKLEAIEAISTESQLEHAKVTQTELM	1858
QY	1387	EOSLNNKEDKDNETTKIVSEM-----	1410
Db	1859	RESFRQKQKATESLKQOEELRERLHEESRAREQAVELSKAEGVIDGYADEKTLFERQIQ	1918
QY	1411	PKDSALLRIEIMGLSKLOESHEDEMSVAKEDDLQRLQOEVLOSSES-----DQLKENIK	1466
Db	1919	EKTDIIRLEQELLCASNRLOELEAQQOIQOEBERELLSROKEMKAAGVPEQOLLOETE	1978
QY	1467	EIVAKHLET-----EELKVAHCCLKEQETINELRVNLSEKETEISTIOKOLE	1515
Db	1979	KLMREKLEVOQAEKVRDDLOKQVKALEIDVEQVSRFIELE--QEKNTLMDLRQOQ	2035
QY	1516	AINDKLQNK-----IQEIKYEBQOLNI-----KQISEVQENVNELKOF	1553
Db	2036	ALEKQLEKMRKFLDEQAIDREHERDVFQOIQLEQOLKVVPFPQPISEHQ--TREVEQL	2093
QY	1554	KEHRAKDSALQSTESKMLELTNLOESQEEI-----QIMIKEEEMKRVQ--	1600
Db	2094	ANHLKEDTKCSELLSKEQLOQDIQERNEEIEKLEFRVRELEQALLVSADTFQKVEDRK	2153
QY	1601	-----ALQIERDOLKENTKEIVAKMKESQEKYQFLKMTAVNETQEKWCE	1645
Db	2154	HFGVAKPELSLEVQLOAERDAIDRKEKIT-----NLEEQLEQFRE-----ELENKNEE	2204
QY	1646	IEHLKEQFETQK-----NLNIEFTENRLTOILHENLEBMSVTKERDRLRSVETL---	1698
Db	2205	VQOLHMQLEIQKKESTTRLOEQEEN--KLFPDDMEKGLAIKESDAMSTQOHHVLF	2260
QY	1699	-----KVERDOLKENL-----RETITRDLEKQEBELKIVHML--	1730
Db	2261	KFAQIIQKEVEIDOLNEQVTKLOOQLKITTDNKBIEKNEILRDLEQIECLMSDQECV	2320
QY	1731	KEHQETIDKLGIVSEKTEINSN-QK-----DLHNSDALKAQDLKIQEB	1775
Db	2321	KRNEEELEQNEVIEKLOQELANIGQKTSMAHLSSEADSLKHQDVLVIAEKALQOQ	2380
QY	1776	LRIAH-----MHKQEQETIDKLGIVSEKTD-KLSNMOKD	1810
Db	2381	VETANEEMTFMKNVLKETNFMNQLTQELFSLKRESESVKIQSPENSNNVAIDHLSKD	2440
QY	1811	LENSNAKLOEKTOELKANEHQ-----LITLKQDVNETQKQVS-----EMEQJK	1853
Db	2441	KPELEVLTIE--DALKSLENTQYFKSPENGKSGIINLETRLLQLESTVSAKDLEITQY	2498
QY	1854	KQIKDOQS-----LTLKLEIENLNTLAQELHENLEBMSVNVK-----	1889
Db	2499	KQIKDOQSGQGPETEMLOKQIVNLOKIVEEKVAAALVSQIQLEAVQYAKFQDNQTISS	2558
QY	1890	--ERDNLRRVEBTLKLRDOLKESLOETKARDLEIQO---ELKTARMLSKKHETVDKLR	1944
Db	2559	EPERTNIQMLNQ--LREDELGSDISALTRLISELESQVEMHTSLILEKEQVEIAEK--	2613
QY	1945	EKISSEKTIQISDIQKDLQSKDELOKQELQKQELQILRVKEDVNMHSHK---INEMBQ	2001
Db	2614	-NVLEKEREKLLBLQKLEGNKKQEKREKRSPOQVEVLKTTTELPHSNEESGFFNELEA	2672
QY	2002	LKKQFEPNLYCKEENDFOLTTKKLHESLEIEIRIVAKE-----RDELRIKES	2048
Db	2673	LRAE---SVATAELASY---KEKAELQE-ELLVKETNMTSLQKDLQSVDRDLHABAK	2725
QY	2049	LK-MERDQFIATLERMIADRQNHQVQKPEKLLSDGQOHLMESLREKSKIKELLKRYSE	2107
Db	2726	LSILEKEDETFVESKCAKCMFPLPTKLKSKSIASQTDGTILKISSNQTFQI--LVNAG-	2782

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:27:31 ; Search time 113.114 Seconds
(without alignments)
7428.155 Million cell updates/sec

Title: US-10-045-631B-88

Perfect score: 13329

Sequence: 1 MAEGAVAVCVVRPLNSRE.....SOPGPHASSGKDVPECKTQ 2663

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 31518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3615.5	27.1	2954	13	O42263	O42263 xenopus lae
2	2234.5	16.8	549	11	Q7TPY4	Q7tpx4 mus musculu
3	1941	14.6	699	11	Q8BWX6	Q8bwx6 mus musculu
4	1911	14.3	699	11	Q80YB4	Q80yb4 mus musculu
5	1562.5	11.7	2244	5	Q9NCGO	Q9ncg0 drosophila
6	1545.5	11.6	2013	5	Q9VKIO	Q9vki0 drosophila
7	1495	11.2	1885	5	Q869B8	Q869b8 dictyosteli
8	1351	10.1	2756	10	Q9LJ60	Q9lj60 arabidopsis
9	1286	9.6	1931	5	Q9VKH9	Q9vkh9 drosophila
10	1279.5	9.6	1931	5	Q9NCF9	Q9ncf9 drosophila
11	1089	8.2	3187	11	Q63714	Q63714 rattus norv
12	1079	8.1	1967	10	Q8GVH3	Q8gvh3 oryza sativ
13	1077.5	8.1	1388	4	Q9NS87	Q9ns87 homo sapien
14	1074	8.1	1463	5	Q9GYZ0	Q9gyz0 strongyloce
15	1056	7.9	1909	5	Q25893	Q25893 plasmodium
16	1048.5	7.9	1388	13	Q91785	Q91785 xenopus lae

Q9w6v0 gallus gall
Q7tnl7 rattus norv
Q8i525 plasmodium
Q7tsp2 rattus norv
Q86kf8 dictyosteli
Q9s7p3 arabidopsis
Q9lut5 arabidopsis
Q94hv9 arabidopsis
Q8ibv8 plasmodium
Q9ul41 leishmania
Q9ngq2 dictyosteli
Q9uf5 homo sapien
Q9lq62 arabidopsis
Q96qg8 homo sapien
Q9v7g9 drosophila
Q8mid9 drosophila
Q86za2 cochiobolu
Q86tn3 homo sapien
Q9n2m3 plasmodium
Q80yp3 mus musculu
Q8i5u0 plasmodium
Q25662 plasmodium
Q8ij44 plasmodium
Q9qz84 mus musculu
Q8x0c5 neurospora
Q7xnll oryza sativ
Q7yue9 plasmodium
Q7ywe8 plasmodium
Q86vh2 homo sapien

ALIGNMENTS

RESULT 1

O42263
ID O42263 PRELIMINARY; PRT; 2954 AA.
AC O42263;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Kinesin-related protein.
GN XCENP-E.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP MEDLINE=98028574; PubMed=9363944;
RA Wood K.W., Sakowicz R., Goldstein L.S., Cleveland D.W.;
RT "CENP-E is a plus end-directed kinetochore motor required for
metaphase chromosome alignment.";
RL Cell 91:357-366(1997).
DR EMBL; AF027728; AAC60300.1; --
DR PIR; T14156; T14156.
DR HSP; P33176; 1BG2.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IPR001752; Kinesin_motor.
DR Pfam; PF00225; Kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
SQ SEQUENCE 2954 AA; 339964 MW; 439804ED0E592679 CRC64;

Query Match 27.1%; Score 3615.5; DB 13; Length 2954;
Best Local Similarity 32.0%; Pred. No. 5.9e-100;
Matches 996; Conservative 572; Mismatches 932; Indels 615; Gaps 86;

475 TLDTSLSEWNPATKLLNOENIESELSRADYDNLVLDYEQURTEKEEMELKKEKNDL 534
 478 TLESLAEVWSSATLLSEENVESELNSLNAQYNDVLDYEQURREDEDLKLKKEKNE 537
 535 DEFEALERTKK 546
 538 ERFELLEQKKK 549

RESULT 3
 Q8BW6 PRELIMINARY; PRT; 699 AA.

AC Q8BW6; 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Hypothetical lipocalin-related protein and Bos/Can/Equ allergen
 DE containing protein.
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Spinal cord;
 RX MEDLINE=22354683; PubMed=12466851;
 RA the FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573 (2002).
 DR EMBL; AK049676; BAC33868.1; --
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR000566; Lipoclin_cytFABP.
 DR PROSITE; PS00213; LIPOCALIN; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 699 AA; 81197 MW; 7392ECC11656C147 CRC64;

Query Match 14.6%; Score 1941; DB 11; Length 699;
 Best Local Similarity 46.2%; Pred. No. 1e-50;
 Matches 427; Conservative 111; Mismatches 120; Indels 266; Gaps 10;

1781 MLKQEQETIDKLGIVSEKTDKLSNMOKDLENSNAKQEKIQELKANEHQILITLKQVN 1840
 1 MNLKEHQETIDRLMETMSEKTEISNMKELENVNMKQEKVQELKTSERQVLLKADAS 60
 1841 ETOKKVSMEQLKQIKQDQSLTSLKLEIENLAQELHENLEEMKSVMERDNLRVET 1900
 61 EA-----KKELKEQGLTSLKIEMENLNAQIHENLEEMKSVRKERDNLKLEI 110
 1901 LKLRDQKESQETVARDLEIQOELKTARMLSKHEKTVDKLRKISIKTIQSDIQKD 1960
 111 LRMRDQLKDN----- 121
 1961 LDKSKDELQKIQELQKELQLLRVKEDVNMHSHKINEMQLKKQEPNLYLCKEMDNFQ 2020
 122 ----- 121
 2021 LTKKLHSELEIRIVAKERDELRIKESLQERDQFIATREMIARDQRNHQ--VKPEK 2078
 122 -----LREAMLKAHQNHETMKCGK 142
 2079 LLSGQOHLMESLREKSRIKELKRYSEMDDHYECLNRLSLDEKE----- 2125
 143 LLCAG-BYCTGRLEKCFRIEKLGRYSEMANDVECLNKVSLDLERTKTKQKELSVTVRT 201
 2126 -----IEFHIMKKLYLVSYVTKIEQHECINKFEMDFID 2162
 202 KLSLPHYQTKEMEKLLTANQRCSELEFHALKRLKYVLSSTARIKEEQHESINKREMAFIQ 261
 2163 EVEKQKELLIKIQHQQDQVPSRLDKLNQWMDLHIELKDFSESEPPSKTBFQ 2222

262 EVEKQNELQIQSLQSYRIPARDIQ-IKLSQMDLHTEMLKDFSENDFITIKTEVQ 320
 2223 VLSNRKEMTOFLEEWLNTFDEKLNKGIOKENDRICOVNNFNRRRIIIMNESTEFER 2282
 321 VLNRRKEITFLGKWLNTLFDTEENLKSTIOKENKSGLVNFFVHSRITAMINESTEFER 380
 2283 SATISKEMEQDLKSLKEKNEKLFQNYQTLKTS---LASGAQNPPTQDNKNPHVTSRATQ 2339
 381 SATRSKDLQYLYKSLKETTEQLSEVYQTLTASQSVHLHPTVQSTPSTRDSEPOAASGAEQ 440
 2340 LITTEKIRLENSLHEAKESAMHKEKIIKMOKELEVNTDIIAKLOAKVHESNCKLEKTE 2399
 441 LTS----- 443
 2400 TIQVLQKVALGAKPYKERTEDLKMVLKIDLEKMKNAKEFEKEISATKATVYQKEVIR 2459
 444 -----NKIAGAVPYKEETEDLKMQLVKSLEKATAKEFDKILSLKATVEHQEEMIR 498
 2460 LLRENLRSSQAQDTSVISEHTDPQPSNKPLTCGGSGIVQNTKALILKSEHIRELEKIS 2519
 499 LLRENLRGHQAQDTSMISQ--DSQLLSKPLTCGGSGIVQSTKALILKSEYKMGSEIS 557
 2520 KLQONEQLIKQKNELLNNQHLSNEVKTWKERTLKEAHKQVTCENSPKSPKVTGTASK 2579
 558 KLQONEQLRKQNNQLSDNSQLSNEVKTWEERTLKKDSTRETTCCENSPKSPKVTGTASK 617
 2580 KKOITPSOCKERNLODPVKESPKSCFFDSRSLPSHPVRYFDNSSLGLCPVQVQAGA 2639
 618 RRONTSQCRANQLODPVKDPSKSWFDFNRKSLPAPHPIRYFDNSSLGLCPVQDD--V 675
 2640 ESDVSPQPMHASSGKDVPECKTQ 2663
 676 ENVEPKTDLCOASLEKDVQCKTQ 699

RESULT 4
 Q80YB4 PRELIMINARY; PRT; 691 AA.

AC Q80YB4; 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Similar to centromere protein E, 312kDa (fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE=Olfactory epithelium;
 RA Strausberg R.; (2003) to the EMBL/GenBank/DBJ databases.
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC049989; AAH49989.1; --
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR000566; Lipoclin_cytFABP.
 DR PROSITE; PS00213; LIPOCALIN; 1.
 FT NON_TER 1
 SQ SEQUENCE 691 AA; 80186 MW; 1D12F180EEC17078 CRC64;

Query Match 14.3%; Score 1911; DB 11; Length 691;
 Best Local Similarity 46.0%; Pred. No. 7.9e-50;
 Matches 421; Conservative 110; Mismatches 119; Indels 266; Gaps 10;

1789 TIDKLGIVSEKTDKLSNMOKDLENSNAKQEKIQELKANEHQILITLKQVNETQKQVSE 1848
 1 TIDRLMETMSEKTEISNMKELENVNMKQEKVQELKTSERQVLLKADASEA----- 54
 1849 MEQLKQIKQDQSLTSLKLEIENLNAQELHENLEEMKSVMERDNLRVETLKLKRDOL 1908
 55 -----KKELKEQGLTSLKIEMENLNAQIHENLEEMKSVRKERDNLKLEIEMERDOL 110
 1909 KESIQETKARDLEIQOELKTARMLSKHEKTVDKLRKISIKTIQSDIQKDLKSKDEL 1968

```
DB 111 KDN----- 113
QY 1969 QXKIQLOKQKQLQLLRKEDVNVSHKKNEMEQKKQFPNLYCKCEMNFQTKKLHSHS 2028
DB 114 ----- 113
QY 2029 LBEIRIVAKERDELRIKESLKMWDQFIATLREMIARDRONHO--VKPEKRLSPGQOH 2086
DB 114 ----- 113
QY 2087 LMSLRKCSRIKELLKRYSEMDDHYECLNRLSLDLKXK----- 2125
DB 142 CTGRLREKCFRIKLLKRYSEMANDYECNLKVSJDLERETTKQKLSVTVRTKLSLPHQTQ 201
QY 2126 -----LEHRIMKLLKYLVSYTKIKBEQHECINKFEMFIDIVEKQKEL 2170
DB 202 TKEMKLLTANQRCLEFHRALKRLKYLSSARIKEQEHESINKREMAFIOEVEKQNEL 261
QY 2171 LIKIQHLOQDCDVPRELRDLKLNQMDLHIEILKDFSEFPSTKTEFQOVLNKRKM 2230
DB 262 QIQISLSQTYRIPARDLQ-IKLSQMDLHIEMLKDFSENDFLTKTEVQOVLNKRKEI 320
QY 2231 TOFLEWLNTRPDIEKLXNGIOKENDRICQVNNFNNRIIAIMNESTEFERSATISKEW 2290
DB 321 TEFLGKWLNTLPDENLAKSTIQENKSGLVNVPYHSRITAMINESTEFERSATRSKOL 380
QY 2291 EODLSLKEKNEKLPKNYOTLKTSS--LASGAQVNPPTQDNKNPHVTSPATQLTTEKIRE 2347
DB 381 DOYLKSLKETTEQLSEVYQTLTASQSVVHLHTVPTQPSRDSRPOAASGAELTS----- 435
QY 2348 LENSILHAKESAMHKSIIKMQKELEVTNDIIAKLOAKVHESNCKLEKTKETIQVLODK 2407
DB 436 -----KNK 438
QY 2408 VALGAKPYKEIEDLKMVLKIDLEKKNKAKEFEKEISATKATVBYQKEVIRLLRENLR 2467
DB 439 IALGAVPYKEIEDLKMVLKIDLEKKNKAKEFEKEISATKATVBYQKEVIRLLRENLR 498
QY 2468 SQQAQDTSVSEHTDPQSNKPLTCGGSGVQNTKALILKSEHRLKESIKLKQKQNEQ 2527
DB 499 HQQAQDTSVSEHTDPQSNKPLTCGGSGVQNTKALILKSEHRLKESIKLKQKQNEQ 557
QY 2528 LIKQKNEILLSNNHLSNWKVTKERTLAKRAHKQVTCENSPKSPKVTCTASKKQITPSQ 2587
DB 558 LRKQNNQLLSNSQSLSEVKTWEERTLARDSYRETTICENSPKSPKVTCTASKKQITPSQ 617
QY 2588 CKERNLQDVPKSPKSCFFDSRSKSLSPHPVRYFNDSSSLGCLCEVQNAAGAEVDSQPG 2647
DB 618 CRAQNLQDVPKSPKSCFFDSRSKSLSPHPVRYFNDSSSLGCLCEVQNAAGAEVDSQPG 2647
QY 2648 PWHASSGKDVPECKTQ 2663
DB 676 LCQASLEKDVSOCKTQ 691
QY Q9NCGO PRELIMINARY; PRT; 2244 AA.
AC Q9NCGO;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DB Kinesin-like kinetochore motor protein CENP-meta.
DB CMT OR CG6392.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
DB NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20351410; PubMed=10893249;
```

```
RA Yucel J.K., Marszalek J.D., McIntosh J.R., Goldstein L.S.B.,
RA Cleveland D.W., Philp A.V.;
RT "CENP-meta, an Essential Kinetochore Kinesin Required for the
RT Maintenance of Metaphase Chromosome Alignment in Drosophila.";
RL J. Cell Biol. 150:1-12(2000).
DR EMBL; AF220353; AAF32355.1; -.
DR HSSP; P33176; 1BG2.
DR Flybase; FBgn0040232; cmet.
DR GO; GO:0005699; C:kinetochore; IDA.
DR GO; GO:0007080; P:mitotic metaphase plate congression; IMP.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
SQ SEQUENCE 2244 AA; 257992 MW; FA6AA3B2A541ADE0 CRC64;

Query Match 11.7%; Score 1562.5; DB 5; Length 2244;
Best Local Similarity 25.0%; Pred. No. 5.4e-39;
Matches 645; Conservative 463; Mismatches 847; Indels 621; Gaps 106;

QY 6 AVAVCVRVPLNSREESLGETAQVYWKI-DNNVIYQVDG-SKSFNFDVRPHGNETTKVY 63
DB 8 SIQVCIKVRPCPGITSL-----WQKEGRSHLADSHAEPPYVDYVDEGASQNEVF 60
QY 64 BEIAAPIIDSAIQYNGTIFAYGQTASGKTYTMMGSEDLGVIPRAIHDFQIKKFFDR 123
DB 61 DRMARHIVHACMQGFGTIFAYGQTSSGKTYTMMGDEQNGVWLAAKEIFQOISSETER 120
QY 124 EFLLRVSYMEIYNEITDILCGTQMKPLIITREDVNRVYVADLTVEVYVYSEM-ALKWI 182
DB 121 DFLRVGVIEIYNEKIYDIL--NKKNDLKHESGNGVNVN--CEECIITSEVDLLRL 176
QY 183 TKGEKSRHYGTVMNORSRSHITFRMILEGREKEGPEKSGSVKSHNLNVDLAGSERA 242
DB 177 CLGKERTVGTNWNERSRSHAFKIIIEGR-KSDHSD-DDAVTQSVNLNVDLAGSERA 234
QY 243 AQTGAAGVRLKEGNCINRSLFILGQVKKLSDGVGGFINTYRDSKLTILQNSLGNPKT 302
DB 235 DOTGARGARLKEGGHINKSLFLSNVKSLENADNRFTNYRDSKLTILQNSLGNPKT 294
QY 303 RIICITIPVSFDELTAQFASTAKYKMTQVYNEVSTDEALLKRYRKEIMDLKQLEEV 362
DB 295 SIICITIPSIMESQSTLSPATRAKIRIKQVNMVSDATWMKRLEREIKVLKDLAE- 353
QY 363 SLETRAQAMEKDQLAQLLEKDLKQVQNEKIENLTM-----LVTSSTSLTQOELKA 415
DB 354 -----EER--KNENQKVEHLERQIKHDMKIKICGHSL-----DKG 388
QY 416 KRKRRTWC-----LGKINKMKNYADQF---NIPTNITTKTKL-----SI 455
DB 389 QOKRRRTWCPTASGSHLEAETGTEDRIDQFPKVSHPKPVFFHTSNAGKRWNPKIT 448
QY 456 NLLREIDBSVCSDESVDNTLDTLSEIENWPATKLNQENIESLNRADYNLDVYE 515
DB 449 NILGSLD--IGTES-----NSIE-EFLPAECI-----DFGSPRPDVLKPLMTIR 490
QY 516 QLRTEKEEMELKKEKNDLDEFEALERKTKKQEMQLIHEISNLKLVKHEVYVNOLEN 575
DB 491 QL-----PDLPTPKGFL-----TTDKIKR-----IQDLQMTSLKHFVEEVEEQG 534
QY 576 ELSSKVELLRKEDQIKKQLQEIYIDSKLENIKMDLSYSLESIEDPKQMKQTLFDAETVAL 635
DB 535 -----LKEKLAEVTAQRDNLEQESL-----AEKERYDALEKEVTS 570
QY 636 DAKRESAPLSENLKELKQKELATTKQKENDIQL---YQSQLEAKK---KMQVD--- 685
DB 571 RADNEAA--NSKISELEKSLTKQTMRIEMEVENQVAVGLEFEFEAKKSKLRVDDL 628
QY 686 --LEKE-----LOSAPNEITK--LTSIDG-----KVPKDLICNL--ELEGIKD 724
DB 629 ALLEKESTIESLQKSLDNLTRDLRNSKSGHMLSIAPQEDVAGDSICNCKEELKIAD 688
```

1702 RDQKLENRETITRDLEKQBELKVHMLKEHOETIDKRGIVSEKTBISNMQDLHS 1761
1615 ANELSNLQKV-----MSLHELIDSQGI-SSDVEINELRELKAA 1657
1762 NDA---LKAQDLKIOBELRIAHMLKEQOETIDK---LRGIVSEKTDKLSNMQ---KD-L 1811
1658 MDKATASAEQMTIVTLQKDVVERMANOAEKFTREANLKGSIINELLKLSNMQETKMDL 1717
1812 ENSNAKLOKIOB-----LKANEHOLITLKDVTNETOKKYSEMEOL---KKOLKQDS 1860
1718 ESGNEBELQOLRNSQNLNMLDESKMCISLKEKLVKLEDAKTSLQOQADNKEIYQRH 1777
1861 LTLKLEIE-NENLAQELHENLEEMKSVMKERNLRRVEETLKLERDQKESLQETKARD 1919
1778 TELTK-EVELGNRIKELTKCEELCSLENSDOIR-----LDIQTKEQLKTKLENN 1829
1920 LEIQOELKTARMLSKHEKTVDKLRKISEKTIQISDIQKDLKSDKDELOKKEOELOKKE 1979
1830 LGWQQ-----KVDVETRECEKLRPFDMQSKVQESKVOEL--- 1864
1980 LQLLRVKEDVNNSHKKINEMEQKKQFEPNYLCKCEMDNFQLTCKLHESLEERIVAKER 2039
1865 -----ISECELRSTL-----KSEASQSEK---ESMD---RTTSSIL 1897
2040 DELLRIKESL-----KMERDOFIATLREMIARDRONHVKPEKELSDGQOHLMESL 2091
1898 EDRNLEELKLSANDIVAKLETE--TAALRPKSLDR---NPVPRKSITFE----- 1943
2092 REKCSRIKELLKRYSEMDDHYECLNRLSLDLEKEIEFHRIMKKLY-----VLS 2140
1944 -----SEIRKN-----RRISVHDERRQSYWVNDVREFGIMTDPVGMSSCVFP 1995
2141 YVTKIE---EHECINKFEMDFIDEVQKELLIKLOHQQCDVPSRELRLDKLQNM 2197
1986 LVFPLNESFRQNC-----NCAELNSKLOQCORELFI---RESQVTLKMWEL 2030
2198 DLHIEETLKDFSESEPSIKTEFQOVLNKRKMTQFLEEWLNTFRDIEKLNGIOKENDR 2257
2031 DHF---PLKO-----ENALTKRVIEEQDKAKVEQKRLKMKLODLNAR 2070
2258 ICQVNNFFNRIIATIMNESTEFEB--RSATIS--KEWEQDLKSLKEK-NEKLFNYQTLK 2312
2071 INDLTT-----ASAKEPESNQAAKAPATVAQOTQESDLETLETKLVNVEEAVMLR 2125
2313 TSLASQAVNPTQDNKPNHVTSPATOLTTKEIRELENSLHEAKESAMHESKIIK 2368
2126 YRYNLIQELKEKURQNSD-TSNITSLAGQTSALKVAI-----SNRYSSSYIK 2175

725 LOKELNKEVEENALREEVILLSELKSLPSEVERLKRKEIQDKSKSELHIITSEKDLFSEV 784
689 LESKNS-----CECDQLEIVSVRDKLESVESAFNLASSGI 726
785 VHKESSVOGLLEIGIKTKDLDATQSNYKSTQBFQ---NFKTLMDFE---OKYKWL 837
727 IQKATDCERLSKELSTSONAFGLQERYDALDOOQAQAGIITLNEHEHVEQEIYOKLQ 786
838 ENERM-----NOEIVNLSKEAKFSSSLGALAKTSLSYKTOELQEKRETVQERLNE 889
787 EYEQLESARSASSAEFORLQNDNTKFOADIASL-----NELEEA 828
890 EOLKEOLERNRDSPLQTVREKTLITIKLOOTLEEVKTLTQEKDDL-KOLQESLQIRRDQ 948
829 QNMLTEVQNSSEIVEKLRIQNHETAKIKELENTPEEMQREYDCLFNQLMESVQ-ENDAL 887
949 KSDIHTVNMNIDTOPQLRNALSRLKQHOTINTLKS--KISEEVENLHMEENTGETKD 1006
888 REBIKQPTSHVSESMRSSGISSDFDEQKQDINLLHQFVQLSEV----- 932
1007 EFOKMGVGDKKODLEAKNTQTLTADVKNEITEQOKFISLQEKNELOOMLESVIAEK 1066
933 -----QOIELQHSGISRLFRANQM-----KLDQSEPLKLCLESA----- 968
1067 EOLKTDLKENIEMTIENQEBELRLIGDELKFKQOEIVAOEKHAIKKEGELSRTCDRLAEVE 1126
969 EYIEEDNRQS-----DATEPICLKGFLKRRHQIKRLSQEHV--DMGEEKRLDIISOLE 1021
1127 EKLKESQOQLOHQOQOQLLVQOEMSEMOKKINEIENLKNELKNELTLEHMEVTERLELAQ 1186
1022 QEIEKSALMEATEATINEMREQNTLESALLEKSVIINKVYQORQIESLEKQNAEMTM 1081
1187 KLENVEEVKS-TYKERKVLKLOKSPETERDHLRG-----YIREIATGLQTKBELKIA 1240
1082 V-----YELOQDVRTRESMESLUR-VPPEDITLPGCPTSPSRQEQVALTKT----- 1129
1241 HIHLKHEQETIDELRRSVSEKTAQIINTOLEKSHTKLOBEIPVLHBEQELLP--NVKKV 1298
1130 -----SITELOSQVSDINALEN-----HLRQLKQDGNLARI 1162
1299 SETQETVNELELLTEQSTTKDSTTLARIEMERLRLNEKFOESOEI-KSLTKERDNLKTI 1357
1163 QTDPEMSERCL-----SMEVRLAELDEDTKQKQLLDQOAKLSDDLCLI 1208
1358 KEALEVQHDQKEHIRETLAKIQSOSQOEOLANMKEDNETTKIVSEMEQFEP----- 1411
1209 -DLOKKNAAQVQVHKATESLSLADAKPDQIILLSSQYDSQIEKLNQLNNAKDELHDVR 1267
1412 --KDSALIRIEIMGLSKRLQESHDENKSVAKKDDQLRLOEVLOSQESDQKENIKETIV 1469
1268 RIKODEISALRMEFL-----LQIETNEKENQAKFYAELQETKDYESNVABLK 1315
1470 AKHLETHEEL-KVAHCCLKQEBETINELRNLVNLSEKEYEISTI-----OKOLEAINDKLQNK 1524
1316 EKLQVETLSSVTVRCQAEALKAHSAHKNISQAVEERNLIIVQOAEEMETIRETLKNG 1375
1525 IQEI---YEKEBQNLNKOISVOENVNE-LKQFKEHRKAKDSALQIESKMLNLTN----- 1576
1376 LAEASTQOSKWEADAFRAINEVRATLMQNLQTKEDRDGKASKLEEVKTKLEQMINGRVR 1435
1577 -----RLOESQEEIQIMKE-----KEEMKRVQBALQIE---RDOLK-----EN 1612
1436 MSDTIAELEKTKAEQDLAVNKLTKDNIELEKQSKTQOLQWESLTRDQISFEIHAHLK 1495
1613 TKEIVAKNK-----ESQEK-EYQFL-----KMTAVNETO-----EKMCIEHLEK----- 1651
1496 LELIVASSKRIIELEEKDDQOQVLELDKCRLEKLSLESIKANESEHCTMEKLOELQAE 1555
1652 -----QFETOKNLNE-NIETENIRLTQILHNLEEMRSVTKERDDLRSVEETLKV---E 1701
1556 MKVLSNRNEKCDPTEKLTETFTFKTD-LEEVLEAKHQKVLVLYDDLSQHERLKICLAE 1614

RESULT 6
Q9VK10 ID Q9VK10 PRELIMINARY; PRT; 2013 AA.
AC Q9VK10;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Cmet protein.
GN Cmet OR CG6392.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

QY 1369 KCHIRETLAKIQESQKOEQSLNWKEDNETTIVSEMEQFKP-----KDSALLRIE 1420
 Db 1219 VEQHYKATESLSIADAKPDQIILSSQVDSQIEKLNQLNNAKDELHVDRIKDEISALR 1278
 QY 1421 IEMGLSKRLQESHDEMSKVAKEKDLQRLQEVLSQESDQLENKEIIVAKHLEBEL- 1479
 Db 1279 MEFL-----LQIETNEKENQAFYAELEQTKDRYESVNAELKEKLQVETELS 1326
 QY 1480 KVAHCCLKEQEBETINELRNLSEKETEISIT-----QKLEAINDKLNKQIEI--YEKE 1532
 Db 1327 SVTVRCQAELEALKSAHAKENISQAVEERNLIVHOQAEOMETIRETLKNKLAESAQOSKM 1386
 QY 1533 EQLNIKOISVOENVNE-LQKFKHRKAKDSALQIESKMLELTN-----RLQES 1581
 Db 1387 EDAFRASINEVRATLMEQLNQTKEDRDGKASKLEEVKKTLEQMINGRVMSDTIAELEKT 1446
 QY 1582 QEEIQIMIKE-----KEEMKRVBALQIE--RDOLK-----ENTKEIVAKWK-- 1621
 Db 1447 KAEQDLAVNKLTKDNIELEKQCSKTQRLQWESLTDQISFEIBAHIKKLELIVASSKXR 1506
 QY 1622 --ESQEK-EYQFL-----KMTAVNETQ-----EKMCETLHLKE-----QFETQ 1656
 Db 1507 IIELEKCDQOVLEDKRLEKJLSLEIQKANSEHSCMEKQLOEAEMKVLNRNEKE 1566
 QY 1657 KLNLE-NIETENIRLTQLENLEEMRSVTKERDRLASVEETLKV---ERDQLENLRET 1712
 Db 1567 KCFETKLEFTFKITD-LEEVLEAKQHVLYDDLVASHQERLAKICLAERANELSNLQKK 1625
 QY 1713 ITRDLEKQEBELKIVHMLKHEQETIDKLRGIVSEKTEINISNMQKLEHSDA---LKAQD 1769
 Db 1626 V-----MSLHTELIDSQGI--SSRDVEINREELKAAMDADAKATASABQ 1668
 QY 1770 LKIQEELRIAHMLKQEQETIDK---LRGIVSEKTDKLSNMO--KD-LENSAKLOEKI 1822
 Db 1669 MTLVTQLKQVEERMANQAEKFTREANLKGINSINELLKINSQETKMLSGNELKEQL 1728
 QY 1823 QE-----LKANEHQILITLKKDVNETQKVSMEQL-----KQIKQDSQSLTSLKLEIE-N 1870
 Db 1729 RNSQNLRLNMLEDSQWCSLSKEKLVLEDAKTSLEQLRDNKNSBIYQSHTELTK-EVELG 1787
 QY 1871 INLAQELHNIIEEMKSVKMDNRRLRVEETUKLERDQKESLQETKARDLEIQELKTAR 1930
 Db 1788 RNRIGELTKKECELCSLENSDQIR-----LDIQTETKEQLAKTLENLQWQ----- 1834
 QY 1931 MLSKEHETVDKLEKISEKTIQISDIQKDLKSKDELQKQIQLQELQKELQLLRVKEDVN 1990
 Db 1835 -----KVDVETRECEKLRFDQSKVEQVNESKQEL----- 1864
 QY 1991 MSHKKINEMEQLKQKQFEPNYLCKCEMDNFQLTKKLHESLEIRIVAKERDELRIKESL- 2049
 Db 1865 -----ISECELRSTL-----KSKEASFQSEK---ESMD--RTISSLLEKRNLEEKLC 1908
 QY 2050 -----KQERDQFATIREMIARDQ-----NHQVKEKELL--SDGQHLMESL 2091
 Db 1909 SANDIVAKLETE--IAALRPKSLDRNPVPRKSITFESEIRKNRISVHDERROSYWNDV 1966
 QY 2092 RE 2093
 Db 1967 RE 1968

RESULT 7
 Q869B8 PRELIMINARY; PRT; 1885 AA.
 AC Q869B8;
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Kinesin-related protein K4.
 GN K4.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycotozoa; Dictyostellida; Dictyostelium.

OX NCHI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98359834; PubMed=9693369;
 RA de Hostos E.L., McCaffrey G., Sugang R., Pierce D.W., Vale R.D.;
 RT "A developmentally regulated kinesin-related motor proteins from
 RL Dictyostelium discoideum";
 RN Mol. Biol. Cell 9:2093-2106 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Suyama E., Sutoh K.;
 RT "Kinesin-related proteins from Dictyostelium";
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB102780; BAC56912.1; -
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005875; C:microtubule associated complex; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003774; F:motor activity; IEA.
 DR InterPro; IPR001752; kinesin_motor.
 DR InterPro; IPR003345; M_repeat.
 DR Pfam; PF00225; kinesin; 1.
 DR Pfam; PF02370; M; 6.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM0129; KISC; 1.
 DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN; 1.
 DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.
 DR SEQUENCE 1885 AA; 218216 MW; 15686AED3B007EC7 CRC64;
 SQ
 Query Match 11.2%; Score 1495; DB 5; Length 1885;
 Best Local Similarity 25.6%; Pred. No. 4.7e-37;
 Matches 544; Conservative 402; Mismatches 708; Indels 470; Gaps 80;
 QY 3 EGVAVCVVRPLNSREESLGETAQVYKTDNNVI-YQVDGSKSFDFRVFHNETHKN 61
 Db 19 ELNKIKVAIRVPLNSRE--LGIDQKIPWSIKDITLSQNPINFTYDYVFGIDSNITD 76
 QY 62 VYEIAPIIDSAIQVNGTIFAYGTASGTYTMGSEDLGVIPRAIHDFQIK-KF 120
 Db 77 VYNAIAKSVNSSLNGITIFAYGQISSGKTFMRGTESIPGIIKLSIKDFSIEDSI 136
 QY 121 PDREFLRVSYMEYINETITDLCGT-QKMKPLIREDVNRVNVADITRVVYVSEMA 179
 Db 137 LEKDYLLKVSYLEIYNEIEKDLLNPTISNKKLKIHEIDYKGVVAVNLKEBIVISPD 196
 QY 180 KWIYTKGSKSRHYGTQKQNRSSRHTIFRMILESREKPECSNCEGSKVSHNLVDLAGS 239
 Db 197 ALMNFGEERRHIGSTMMNDSSRSHITFRMQIQSTCKQ-----NGTIQMSTLTLDLAGS 251
 QY 240 ERAQGTGAAGVRLKEGNNINSLFTLGQVIKKLSQGVGGFINVRDSDLTRILQNSLGN 299
 Db 252 ERVSTGAEGVRLKEGTHINKSLMTLSKVISLSEKTKQOHVPYEDSKLTRILOPSLGN 311
 QY 300 PKTRICITTPVSF--DETLALQFASAKYKNTTPYVNEVSTDEALLKRYRKEIMDLKK 357
 Db 312 SKTALLCITITATTHQEEISITLQAKRAKRVKTKYKINQVADANTMLKYESEILEQN 371
 QY 358 QLEEVSLTEAQAEMKQDLAQLLEKOLLQVQEKIENITRMVTSSSL----- 407
 Db 372 QL-----VKSEINSLRNTISTQEISSNNFKLGMK 402
 QY 408 -----TLQQLKAKRKRVTW-----CLGKINKMKNISYADOFNIPNTITT 448
 Db 403 RFDALIGSLINENKKKKKRNTLDPSYLLKDKLIIKKIRSENQKIKKINSENNISS 462
 QY 449 KTHKLSINLLREIDESVCSDFVFSNTLDTSEIENPATKLLNQNTESELSNRADYD 508
 Db 463 S-----SSNSGEEDDDKDDENNYISINQDDKADSHY-----EDDDDEDEDDDESDTD 511
 QY 509 N-----LV--LDYEOALRTEKEEMELKLEKKNLDLDEFEALE 541
 Db 512 NEDDEDNEDNDDDDDDDEFQDNPLPLPLEDDQLK-----KIK---DUDSLGFQ 560
 QY 542 RRTK-KQEMQLIH-EISNLKLVKHREVINQDLNELSSKVELLEKEDQIKKLQFYTD 599

Best Local Similarity 22.7%; Pred. No. 1.3e-32;
Matches 626; Conservative 513; Mismatches 1020; Indels 598; Gaps 102;

7	VAVCVVRPLRSRESL-GETAQVYWKTDNNVYIQVDSKSFNDRVYFHGNETTKNVEE	65
194	VQILIRVPLNSMERSINGYARCLKQSSQCVAMIGPPETRFQPDHVACETIDQETLFRV	253
66	IAAPIIDSAIGYNGTIFAYGOTASGKTYTWMSSEDL-----GVIPRAIHIDFQKI	117
254	AGLPVENCISYNGICIFAYGOTGSGKTYTWLGEVGDLEFKPSPNRMGMPRIEFLFARI	313
118	KKFPDR-----EFLRVSYMEIYNETTIDLCGTQGMKPLIREDVNRNVYVADLTREV	171
314	QAEESRRDERLYKNCKSCFLIYNEQITDIL--EPSSTNLQREDIKSGVYVENLCE	371
172	VYTSEMALKWITKEGSRHYGTETKQNSRSHTIFRMILES-RKGEPSNCEGSSVKVSH	230
372	VQSQVDILGLITQGSINRRVGTATNNRSSHSVFTCVIBSRMEKOSTAN----MPAR	427
231	LNLVDLAGSRAAQTGAAGVRLKGCNTNRSFILGQVKKLSQGVG--GFINRYSKL	288
428	LNLVDLAGSERQKTSAGEDRLKEAASINKSLTSLGHVIMVVDVANGKPRHIPYRDSRL	487
289	TRILQNSLGNPKTRIICITTP-VS-FDRTLALQFASATYKMKVTPYNEVSTDEALLK	346
488	TFLLQDLSGNSKTMIIANASPSVCAETLNTLPAQRAKLIQNNAVNEDSDNEDVL--	545
347	RYKEMDMKOLLEVSLETRAQMEKDOI-----AQLLEKDLQVQNEKIENL---	397
546	ELRRQILLKEELSLKRONISRALFSFGSATANFAESQVDSPPSSVHTTGQOAGNLLVY	605
398	-----TMLVTSSSLTLOQLKAKRK-----RRVTWCLGINKNMKNSYADQFNPT	444
606	ESGCVMSRKQKSLIETLAGSLREHVADASIKLEABIEHLNRLVQREED-----	659
445	NITTKTKLSINLRE-----IDESVCSDDVPSNTLDTI-SEIENPATK	489
660	---TRSTKMLRFREDKIORLESLLGNHSADSFLEENNVLSEETQLQAKIDNPBLT	716
490	LMQENIESEINSLRADYNLDVYQART-----EKEEMELKKE-----KNDLDEF---	537
717	RFALENIR-----LLD--OLRRFOFYEEGREILLGVEVSLNRNLQFLDE	761
538	-----BALERKTKK-----DOEMOLIHIEISNLKVLKRVHREYVNDQ	572
762	NSDWQKVDGIFPOLKTKCYELEKCSNLGSCLEENAKLSREINDLQAMVSDIRACTPD	821
573	LENELSKVLELREKE-----DOIKKLOEYDSOKL-----EN	605
822	EHSVNVKQALLGQNPETHETLACEQANYVEEIIQLDLDVQKIILDEBTLRGDTBA	881
606	IKMDLSVLSIED-----PKMQKT--LPDAETVALDAKRESAFLRSENLELKEKKE	657
882	QAVRLKFDIEVLQDOLLISQOQNVYSELGETKSAVALESQNIILQAEVLELRIKEN	941
658	LATTYKQMDNIQLYQ-----SOLEAK-KMQQVDLKELESQAFNEITKPLTS	702
942	YFELLKQELDIPAKSKQCDKPNFAEDSEIDTFKKMQASLEKAKRLNMLYKSDIAS	1001
703	LIDGKVPKOLLN-----LELGGKTDLOKELN-----KEVENEALREE	742
1002	KACGEEMDEVCKQAAATAEVIQCLQNEVLEQVEYNDPQSKENVTEKQVILETQMBE	1061
743	VILLSLSLSPSEVERLKEITQDSKEELHIITSSKOKLFEV-----VHKE	788
1062	--LQDKLRTMDNEQLOEQRGKQMDLIIISNEMELLTSELBEILLNGEGLTDACYQA	1119
789	SRVQGLL-----EETGKTKDGLATTQSNYKSTQDFQNFKTLHMDPFQKYKMWLEEN	840
1120	DLISGLDPDKRIWISQVGGIIRTLSERELMIEDLESLEDANKKRCDIESMLKSLKGAA	1179
841	ERMNOEIVNLSKEAQFDSSIGALKTELUSYTKQ--FLOEKTRVQERL-----	886

QY 484 WNPATKLLNOENIESELNSRADYDNLVDYERLQTEKEEMELKUKKXNDLDFEALREK 543
 Db 409 LRPSPKLTNLPFLPQRRGIAPKAGICKTLKELQTDNMD----- 450
 QY 544 TKQDQEMQLIHEISLNKLVKREVNQDLEN-----ELSKV-----ELLREKEDQ 590
 Db 451 TPGRAKQLGRETASRIPSWMSKKYQESVPNCAPQTEISALTASNOVAKETIEKYBEQ 510
 QY 591 IKKLQYIDSOXLENIKMDLSYLSIESIEDPKQKQTLFPAETVALDAKRESAFLSENLE 650
 Db 511 VRLKRETBELREMGKA-----VN 530
 QY 651 LKEKMKELATTVKQMDNDIQLYQSOLEAKKQMDVLEKELQSAFNBITKLSLIDQKVPK 710
 Db 531 LGEOPETHKAKSKQMEEL---LSSISEKSTIVSLQSLLE-----ELSR 572
 QY 711 DLLCNLELGGKITDLOKELNKEVEE---NEALREVIL---LSELSKLPSEVERLKEIQ 764
 Db 573 DVLRSK-EDQMSMCPLESSCERICNCKLELERLLPLASAGLDSVACQFQDLRSEIA 631
 QY 765 DKSEELHIITSEKDLFSEVVHRESVQGLLEIGTKKDLATVQSNYKSTDOEFQNFK- 823
 Db 632 ATMKLESMLSTFSHASCESVQKTDCKRLSEQISTAHDDFGQLQEKYNNLKHKWSQKL 691
 QY 824 ---TLHMF---PQYKMWLEENE---RMNQEVNLSKEAQFSDSLGALKTELKSVKTQ 873
 Db 692 AIDTMQVDYNTIOQYLOQDEYRHLELRSDEQOQLODENSKLQABEIGTLKE---RVE 747
 QY 874 ELQEKTRV---OERLNEMOLKELENRSPLO---TVREKTLITEKLOQLTLEVK 926
 Db 748 EIHSELLEVPNPHDEMLQONQELKRLSKLQWEPDEIOLNVECSNLMSTIQECDA 807
 QY 927 LTQE-----KDLQLOQESLIQIERDOLKSDIHDVTN-MWIDTOEQLRNALESILKQ--- 975
 Db 808 LREBKORTNSDESMSKSS-----GVGTCSDPENELTDLLQOQTKLSKSQIQLT 861
 QY 976 -----HOETINTLKS-KISEVSRNLHMEENTGETKDFQOKVMGIDKQDLE 1022
 Db 862 DYSGRRLFIYNHAEQDQSVPSLKLCLPAKYL---EGDQKHQDASDVFL---KGLK 914
 QY 1023 AKNTQTLTADVKNENIIEOQRK---IFSLQEKNELOOMLESVIAEKEQLKTDLENLEM 1079
 Db 915 QRFQIVKINGEQLNVEEDRMDIIFQLKQEVDPGKNLIE---EKEVINNNRAQITS 970
 QY 1080 -----TLENQ-ELRLAGDLKQOEIVA---QSKNHAIKKEGELSRTCDRLAEVEEK 1130
 Db 971 LNOETIKQNAKTKILCEEELQTDVOTANKQESQEVLTIKTSLAHLKSVCELOKKLE 1030
 QY 1131 EKS---QOLQEQOQLNVOEEMSEMOKKINIEMLK-----NELKNKELTLEHMET-- 1179
 Db 1031 KQSEDEKISELQSDIGHISECCLSMELKLADIVNWAEBELRPLDQLOESGVELOHHSTA 1090
 QY 1180 -ERLELAQNLNENVEKSIYKERVILKELOKSPETERDHLR-----GYI 1223
 Db 1091 BESLNVEKPIQEQERTLTITYERRI-BQLESIQRAQOEELSIKRTKTDENKSLQLYM 1149
 QY 1224 REIATGLQTEELKIAHILHQETIDELRRSVSEKTAQIINTQDLESKSHLQEBIP 1283
 Db 1150 AKIETSENENSKFRAYCLDLKETQRYEBEOLQOTNEKLAHV-----TQCVHLD 1200
 QY 1284 VLHBEQELLPNVKKVSETQETMNELELITQESTTKDSTTLARIEMRLR---LNEKFSQS 1341
 Db 1201 VIKRSLQ-----EKITQAEKERNEL-----AVRHKAELKIRETLKEKESKY 1243
 QY 1342 BEIKSLTBERDNLTKIKALEVKHQDLKHEIRETLAKIOESQOESQOSINMKEKDNETTK 1401
 Db 1244 EKLRQAEERD---KEISR-LEV-----MENTAELUKINSREVEL-----EGVK 1285
 QY 1402 IVSEMEQFQPK---DSALIRIEIMGLSKRLQESHDEKMSVAKK-DDLQRL----- 1450
 Db 1286 -----MEKQKLYDKSML--ELEQLOCTAD-QKSSDLLPGSSNENIDDLQKKQCVQD 1338
 QY 1451 QEVLOSQESQDKENIKIVAKH-----LETEELKVAHCLKEQBETIN 1494

Db 1339 LELLGEKAEALLSELOKINGQHSNTIKKLEETEAEMITLTQKLEL--RCEIAEKLTFK 1396
 QY 1495 ELRVNLSKETETISTIQOLEA-----INDKLNKIQEIVYEKEBQ 1534
 Db 1397 SKEADIKK---ALHCAQLRLHAYDKLVCEYERLKGCLSDNSKNLSLQKKVERLHQAQLA 1453
 QY 1535 L-----NIKOI-SEVOENVNELKQPKHRKAKDSALQSTESKMLELTNRLOL--- 1580
 Db 1454 LOEGISGRDSEIKQRLSELKDAIDENKTVREAKVGLNSLKAVQENWSAQEGQFKQKAD 1513
 QY 1581 ---SOBEIQIMIKKEEMKRVQEQALQIERDQDKENTKEIVAKWSQOEYQFLKMTAVN 1637
 Db 1514 IKGSVDDELQIKLSLQEV-----RDHLESNEELKRLKDAQELQNMVDKERKLN 1563
 QY 1638 ETQEKMCWEIHLKEQF---ETOKLNLNETETENIRLTOI-LHENLEEMSVTKERDRLS 1693
 Db 1564 SS-----LREDPDKLEOTKLDLE---BOLRAKVEIDRRSKELGEVTKDCENIRS 1610
 QY 1694 -----VEETLKVERDQK---ENLRETITRDLEKQEBELKIVHMLKHEQETIDK 1739
 Db 1611 DLBAQNDPFLKRETLNLITISDLRLHNEQLLETSKNLSDTAANNLNMKKNLHDLTK 1670
 QY 1740 ---LRGIVSEKTEINSMQKOLEHNDALKAQDLKIQEBELRAHMLKQEQETIDKLRG 1795
 Db 1671 ECKSLRSDLSQSEYFQTKQLLDETISNLKEENRWEEKLSSGNKALKDCE---KLRS 1727
 QY 1796 IVSEKTDKLSNMQKOLENSNAKLOK-----IQELKANEHQILITLKK-----DVNE 1841
 Db 1728 TLESKELILOONKQOEELRTVINEKNGKNALLDAQLKSNETAPTSLRKAWIKQSLAIEA 1787
 QY 1842 TQKVSMEQO-LAKOIKDQSLTSLKLEIENLNAQELHENLEEMKSVMKERDNLRR--- 1896
 Db 1788 ANKRSLEMEQDKRTREYEELSTLKTREINFRSEKERMDGTITSSLEDKRNELEKLT 1847
 QY 1897 VEETL-KLERDQKESIQETKARDLETQOELKFTARMLSKHEKTVDKLRKISEKTIQIS 1955
 Db 1848 VTELLAKLKEELPALHTQKVGSDVSI--ELNNG-----SPTPAAPV 1889
 QY 1956 DIQKDLQ-KSDELOKKIQELOKKEQLLQVKEVDNVMHKKINEMEQKKQF 2006
 Db 1890 ATKPLDCAECPVKSSSLTAEL-----RKNRMTAYDENRKQF 1930
 RESULT 10
 Q9NCF9 PRELIMINARY; PRT; 1931 AA.
 ID Q9NCF9
 AC Q9NCF9
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Mitotic kinesin-like motor protein CENP-ana.
 GN CANA OR CQ4831.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20351410; PubMed=10893249;
 RA Yucel J.K., Marszalek J.D., McIntosh J.R., Goldstein L.S.B.,
 RA Cleveland D.W., Philip A.V.;
 RT "CENP-meta, an Essential Kinetochores Kinesin Required for the
 RT Maintenance of Metaphase Chromosome Alignment in Drosophila.";
 RL J. Cell Biol. 150:1-12(2000).
 DR EMBL; AF20354; AAF32356.1; .
 DR HSSP; P56536; 2KIN.
 DR FlyBase; FBgn0040233; cana.
 DR GO; GO:000090; P:mitotic anaphase; IMP.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.

Db	1729	LESKELIIQONQLEERLTVINEKNGKALLDAQLKSN--ETAFKSLQKAWIKOSLAIE	1786	QY	166	DLTEBVVVTSEMAKWIITKGEK-----SRHYG-----ETKMNQRSS	201
QY	1928	TARMLSEKHETVDK-----LRKISEKTTQISDIQKDLKSKDELQKKIQELQKK--	1978	Db	278	ERQESKILMEKVELEMAERKEELYLOQLREAGQAQAELEMOYGTLOQRHETEMEKK--	335
Db	1787	AANKSEMEQWDRTRREYELRSTLTREINFRSEKERMDDGTISSLLDKRNLEEKLC	1846	QY	202	RSHTITFRMILESREKGEPSNCEGSKVSHNLNVLGASERAAQTAAGAVRLKEGCNINRS	261
QY	1979	---ELQLLRVKEDVNMVSH-KKIN	1997	Db	336	---TACISLQKNEQELQACD-ALKEENSKLQ-EOQEQAQAKAQAQLOQLEDE-----	384
Db	1847	TVTEL-LAKIKRELPAHTQKVN	1868	QY	262	LFILQVITKULSDQGVGFINYRD-SKULTRILONSLGGNPKTRIICTIPVFSFDELTAL	320
RESULT 11				Db	385	---LOQKSKEISQ-----FVNKPNEKHETSSQTSU-----PDVNEGVQAV	423
Q63714	PRELIMINARY;	PRT; 3187 AA.		QY	321	OPASTAKYMNTPYVNEVSTDEALLKRVKEMTMDLKQLEEVSS-----LETRAQAMEKQD	375
AC	Q63714;			Db	424	MEESVASLOQR---VLELENEKALLSLELEELRAENEKLCRSRITLLEAQNURAGEADG	480
DT	01-NOV-1996	(TREMELrel. 01, Created)		QY	376	LA-----QLEEE--KDLQKQVONEKIENITRLMVTSSSLT-----LOQEL	413
DT	01-NOV-1996	(TREMELrel. 01, Last sequence update)		Db	481	MVCEVSTAGIALNRSDSSTESGQDVLENTFSQHKELSVLLVEMKEBAQEAETAFKLSQL	540
DT	01-JUN-2003	(TREMELrel. 24, Last annotation update)		QY	414	KAKRR-----RVTWCLGKINKMKNS-----NYADQF	440
DE	GIANTIN	(Golgi complex-associated protein of 364 kDa) (GCP364).		Db	541	QCKRPGDYEVLDKREKVQMMSEGLPSVTARDVLC---APROKNSYPAVEGEQAGMRDQH	597
OS	Rattus norvegicus (Rat).			QY	441	NI-----PTNIT-----TKTHKL-----SINLLREI	461
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			Db	598	GILEAGPLNDTGMLNSPQDGVKSLSAPHVCLCHQGLERLKTQVLETSHTAKET	657
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			QY	462	DESVCSESVFNLTDLTSEIENWNPATKLNQ--ENIESELNLRADYDNLV-----LD--	513
ON	NCBI_TaxID=10116;			Db	658	HEKNLSEK--AKEISSLAQL-----TKFKESAEEARSTLTAVCEERDQLLYRVKELDV	709
RP	SEQUENCE FROM N.A.			QY	514	YEQRTKESEMELKKEKNDLDEFEALERKTKDQEMQLIHEISNLKLVKREVYNQDL	573
KC	TISSUE=KIDNEY;			Db	710	LGLERQAVRELETSLAE---AEKQGLDYESQAQHNLLTEQIHSLSLEAKSKDVKTIEL	766
RX	MEDLINE=980393490; PubMed=9431462;			QY	574	ENELSSKVELLEKEQIKKIQEYIDSQKLENIMKDSLSLESIEDPKQMKQTLFDAETV	633
RA	Toki C., Misumi Y., Fujiwara T., Sohda M., Nishioka M., Ikehara Y.;			Db	767	QRELDGVQLOFSEGTQIKSLQSLQTKQE-----SEVLEGAERMK-----	806
RT	"Identification and characterization of rat 364-kDa Golgi-associated			QY	634	ALDAKRESAFLRSENLELKEKMKELATTYQKQMEMDI-QLYQSOLEAKKKQVLDLEKLOS	692
RT	protein recognized by autoantibodies from a patient with rheumatoid			Db	807	-----DISKEMBELSQAQLEIAQMDQLLELKKDDVET-LQQTQIE	849
RL	arthritis."			QY	693	AFNEITKLTSLIDCKVPKDLLCNLELGGKITDQKELNKEV-----EENEALREEVILLSE	748
CC	Cell Struct. Funct. 22:565-577(1997).			Db	850	KDQOQVT-----ELSFMTKRVQVNEEKFSGLVETKLTKEQLSLLSR	891
CC	-1- FUNCTION: MAY BE INVOLVED IN THE FORMATION AND/OR MAINTENANCE OF			QY	749	L-----KSLPSEVERLRKEIQDKSEELHIITSE	776
CC	THE CHARACTERISTIC GOLGI STRUCTURE.			Db	892	ARGAKRQVEDSGAESPKHGPHESSAEEPVCKEALQOLEWLKESQERKRKLAALIS	951
CC	-1- SUBUNIT: DISULFIDE-LINKED HOMODIMER.			QY	777	KDLFSEVHVHKEVRVQGLLEIGHIKTKDDLATQSNYKSTDOEFONFKTLHMDFOKY---	833
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. GOLGI MEMBRANE.			Db	952	RKELLQKVSLEELAKVREE--STKDSL--RESEKRELEEDSKN-----KDDPEKYGTS	1002
DR	EMBL; D25543; BAA05026.1; --			QY	834	-----KMWLEENRMQEIIVNLSKEAQKPFSSILGALKATSLSYKTOELQKRETV--	882
DR	PIR; JC5837; JC5837.			Db	1003	EWELEVSRLTITSEKEVELEGIRDLKKAABEELQALVQM---TQDLQNKTKIDL	1059
DR	GO; GO:0005794; C:Golgi apparatus; IEA.			QY	883	QERLNM-----POLKEQLNRRSPLQTV	906
DR	GO; GO:0016021; C:integral to membrane; IEA.			Db	1060	LQEHITENQATIQKFIPTGMDAGDGSVAKETSVSSPPRAGGGEHWPKELEGK-----IVDL	1116
DR	GO; GO:0005871; C:kinesin complex; IEA.			QY	907	EREKTLITEKLOQTLLEVKTL---TOEKDDLKQLESIQIERDQKSIDHTVNNNDITQ	963
KW	Golgi stack; Antigen; Coiled coil; Transmembrane.			Db	1117	EKEKTLQKQLEALISRKALKKAQEK--KHLKEELKEQDAYR-HIQEFDQCSKEN	1173
FT	DOMAIN 1 3163			QY	964	EQLRNALSKHOHQTINTLAKSKISEV---SRNLHME-----ENTGETK	1005
FT	TRANSNEM 3164 3184			Db	1174	ENIRAPLQQAQKSTDOQLPQTGOQEPHSGEGSLSEGTREPASSEDHAAQSPHPGETA	1233
FT	DOMAIN 3185 3187			QY	1006	D-----EFQKQVMGID-KKQDLLEAKNTQTLTADV--KONEIIEQKIFSLQEKNE	1054
FT	DOMAIN 47 116						
FT	DOMAIN 126 398						
FT	DOMAIN 418 479						
FT	DOMAIN 505 544						
FT	DOMAIN 632 891						
FT	DOMAIN 924 993						
FT	DOMAIN 1010 1076						
FT	DOMAIN 1103 1178						
FT	DOMAIN 1231 1327						
FT	DOMAIN 1348 1712						
FT	DOMAIN 1765 2320						
FT	DOMAIN 2331 2706						
FT	DOMAIN 2356 2359						
FT	DOMAIN 2723 2775						
FT	DOMAIN 2800 2892						
FT	DOMAIN 2953 3008						
FT	DOMAIN 3069 3113						
SQ	SEQUENCE 3187 AA; 364295 MW; 48F035DF43647F51 CRC64;						
Query Match	8.2%; Score 1089; DB 11; Length 3187;						
Best Local Similarity	20.9%; Pred. No. 9.1e-25;						
Matches 634;	Conservative 558; Mismatches 1012; Indels 828; Gaps 124;						
QY	113	IFQKIKKFPDRFELLRVSVMYVNETIT-----DLLCGTQKMKPLIIRDVNRNVYA	165				
Db	221	VNQRKLEHEEALLGRAQVVDLLQKELISAEQRNQDL---SQQLQLLEAEHSTLNTMEA	277				

Db 1234 TQATVSVAGIQDQLKBIIEVEKELEKISST-TSELTKKSEVLLQEQINEQGLEIQN 1292
QY 1055 LQOMLESVIAEKEQTLDKENIEMTIENQOEELRLIGDELKKQOETVAQEKHAIKKEGE 1114
Db 1293 LKAASHEAKAHTQELQOEL-ESSQLKADLEHLKTLQPELETLOKHVGQKEEVSVLVGQ 1351
QY 1115 LSTCDRLAEVEKKEKQSQ-----QLEKQOQOQLNVQERMSQKINBIENL 1163
Db 1352 LGEKEQTLTVQTEMEQERLIKALHTQLEMQAHEERIKQVQVEICLKQKPELE-- 1409
QY 1164 KNELNKELTLEHMETERLELAQKLNNVEVKSITKERVKLKQKSP-----E 1213
Db 1410 -BESAKOOLQKQOALISRSKEALKENKSLQFOLSSARDAVEHLTKSLADVESQVSVQN 1468
QY 1214 TERDLRG-----YIREIATGLQTK-----EELKIAHHLKHEOETIDELAR 1256
Db 1469 QEKDALLGKALLQBERDKLIVEMQKSLLENQSLGSCESLKLALGGUTEDKEKL--MKE 1526
QY 1257 SVSEKTAQIINTQDLEKSHTKLORIPVL-----HEEQELLPN 1294
Db 1527 LESVRSKIAESTEWQEKHEKQELQEVVLQSYENVSNBAERIQHVESVRQEKQEVAK 1586
QY 1295 V-----KKVSETQETWNELE----- 1309
Db 1587 LRSABSDKREKQDQDAEQEMEMKWKPKAKSKQKILEBENDRLRAEAQPVGA 1646
QY 1310 -----LATEOSTTKDSTTLARIEMERLRLNFKFOESQEBEIKSLTKERDNLKTIKEALEV 1363
Db 1647 NESMEALLSSNASLKEE--LERITLEYTKLSEFALMAEKNTLSEETRNKLQVQAEQEL 1704
QY 1364 KH-----DOLKEHTRE--TLAKIQESQSQKQESQSLNMKEDNETTKIVSEMEQFKPKD 1413
Db 1705 KOASLETTEKSDPKVDVIEEVTEAVVGKSEQDSLSSENAKLEDABATILANSA---KP-- 1759
QY 1414 SALLRIETEMGLSKRLQESDEMKSVAKEKDQLO-RLQEV-LQESD-----OLKENIKK 1467
Db 1760 -----GVSETP--SSHDDINNYLQDQKGRITAELEMEKQDKDRLSQTLENEKN 1807
QY 1468 IVAKHLETER-ELKVAHCLKEQBETINELRVNLSEKETEISTQKQLEALNDKLQNKIQ 1526
Db 1808 ALLTQISAKDSELK-----LEEVAKINMLNQIQEELSRVTKLTAEERKDDLEERL- 1862
QY 1527 EYIEKEBQNLNKIQSEVOENVNELQFKHEHAKDSALQSIESTKMLEITNRLQESQEBEQ 1586
Db 1863 -----MNQALANGSIGNYQVDTDAIKN---EQLESEMQLKRCVSELEBEKQ 1909
QY 1587 IMIKE-----KEMKRVQEA-----LQIERDQKENTKEIVAKMKSQEKQYQFLK 1632
Db 1910 QLVKEKTVSEIKETWEIKQGAQKQPGSKIHAKELQELKKEKQEVKQIQKDCIRYL 1969
QY 1633 MTAVNETQEKQCE-----IEHLKE--QFETOKNLNLENTENIR 1669
Db 1970 RISALEKTVKALEPVHTESQKDLATKGNLAQAVEHHKKAQELSSFKILLDDTQSEAR 2029
QY 1670 LTQILHNLSEMSVTKERDRLSVETLKVVERDQKLENRETITRDLKQELKIVHMH 1729
Db 2030 ---VLADNL-----KLKELQSNKESIK---SQIKQK-DEDLRLRLQAE--KH 2070
QY 1730 LKEHOETIDKLRGVSEKTEWEISNMQKDLSEHNDALKAQDLKQOEBELRIAHMLKQOET 1789
Db 2071 RKEKNQKQELDLHREKA-----HVEDTL-----AEIQVSLTRKDKDKMKELOQS 2115
QY 1790 IDKLRGVSEKTKLSNMQKDLN-----SNAK-----LOEKIOELKANEHLITLTKD 1838
Db 2116 LDSTLAQLAFTKSMSSLODRDRVIDEAKWQRFQDALTQTEEBEVRLEKNECTALK-- 2173
QY 1839 VNETQKVSMEQKQIKQKQDQSLTKLEIENLQAELEN--LEBMSVWKERNLNR 1896
Db 2174 -DQLRQWTHMEELK-----ITVSRLEHD-----KEIWESKAQTELQHQKAYDKLQ 2220
QY 1897 VBETKLKRLDQKESQETKARDELTQOELKTARMLSKHEKTVVDKLR-----KI 1947

Db 2221 ENKELMSQLEAEAGOLYHDSKNELTKLESELKSLKDQSTDLKNSLEKCRHENHNLGIIQ 2280
QY 1948 SKTIQ-----ISDIQKDLKSKD---ELQKTIQELQKELQLLRVKE-----DVNM 1991
Db 2281 QEADIQNCXFCNEQELTDLTASRELTTLRHDEINVKEQKIISLLSGKEAIVQVIAELHQ 2340
QY 1992 SH-KKINMEQLKKQPEPNYLCKCEMDNFQTKLHSHESLEIRIVAKER-----DELR 2043
Db 2341 QHSKEIKELLENLSQEEENL--TLEENKRAVEKTNQLTAELETIKKESIQEQAQLDSFV 2399
QY 2044 RIKESIKMERDQFIATLRMIARDRQNHQVQKPEKLLSD--GOQHLMESIRKCSIKEL 2101
Db 2400 KSMSSLODRDRIVSDYQL-----BERHLSVILEKDELITQDAARENKKEE 2447
QY 2102 LKRYSEMDDHYECLNRLSLDLLEKE-IEFHRIMKGL-----KYVLSVYVTKIKEQHE 2151
Db 2448 IR--GLRGHMDLINSNAKLDAELIQVRRDLNEVITIKDSQORQLLELAQLOQNKELRNE 2504
QY 2152 CINKPEMDFIDBEVQKELLIKIQLHQDCDVPSELRLDLKLNQMDLHIEILKDFSES 2211
Db 2505 CV-KLEGRLKGEAEKQSLQMSLDALOBENQGLSKEIKSFK-EQLTALHEEGALAVY-HA 2561
QY 2212 EPPSIKTEBQOVL-----SNRKEMTQFLEMLNTRFDIEKLKNGIQKENDRICQVANNFNN 2267
Db 2562 QLRVREBEVQKLTAAALSSQKRTVDLQEBELVCVQKEASKKSEIEDKLK--ELKHLHN 2619
QY 2268 RIILMNSTEFERSATISK---FWEQDLKSLKEKNEKLFKNYQTLTKTSLASGAQVNP 2324
Db 2620 AGI-MRNETAEERVAELARDLVEMEQLLTVTKENKDLTAQIAQFGKSS----- 2671
QY 2325 TODKNKPHVTSRAQTLTEKIRELENSHAEKESAMHESKIIKMKQELVTVNDIIAKLQ 2384
Db 2672 LQDSRD-HATELSDL-----KKCYDASLAKELAQKRGQDLGRESQVLSQA 2717
QY 2385 AKVHESNKLEKTKETQVLQDQKVALGAKPKYKEIEDLKMVLKIDLEKMKNAKEFEKE- 2443
Db 2718 FPL-----TTSNI-----SSLEKLNQOLISKDEQLLHLSSESSH 2755
QY 2444 ---ISATKATVEYQKQEVIRLLR--ENLRRSQQAQTVTSEHTDPSPNKELTCGGSGI 2498
Db 2756 NQVQSFQKMTSLQNERDHLWNELEKFKSEEGQRSAA-----PSAASSPAE----- 2803
QY 2499 VQNTKALI--LKSEHIREKESIKLQKQNEQLIKQKNEL-----LSNNQHLSEVNTWK 2550
Db 2804 VQSLKAMSSLODRDLKELKNLOQOYLQMNQOETELRPLKAQLOESQDTKALQMMK 2863
QY 2551 ER-----TLKREAHQVTCENGPKSPKVTGTASKKQITPSCCKERNLQDPV 2597
Db 2864 EELRQENLSQWHELDQLRVEKNSWELHERRMKQYLMASDKQOQGLHQLNLRREL----- 2919
QY 2598 PKESPKSCFPDSRSKSLPSPHPVRYFNSSLG 2629
Db 2920 -----RSSQTQILPTQYQORQASSG 2939

RESULT 12

Q8GVH3 PRELIMINARY; PRT; 1967 AA.
ID Q8GVH3
AC Q8GVH3;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE QJ1340-C08.3 protein.
GN QJ1340-C08.3.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_Taxid:39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Katayose Y.;

[illegible]

[illegible]

1826 SBERLSEKIKKVKYKT-----KDNFKPNDKS-----LYDEHIKY 1864

2516 KEISKLOQNEOLIKO-----KNELLSNNQHLNNEV 2546
1865 KNDKQVNKEKEFKISLPHIFDGDNEILQIVDELSEDI 1902

Search completed: July 29, 2004, 09:39:38
Job time : 134.114 secs

993 ERR-KEKLOEQSDLEQERRAKEKLOEQSDLEQERLAKELQEQSDLEQERLAKEL- 990
1479 LKVAHCCLEQEBETINELAVNISEKETETSTTQKLEANDKLNKIQIYIEKEEOLNIX 1538
991 -----LQEQSDLEQERLAKELQEQSDLEQERLA-KEKLOEQSDL--EQERLAKEL 1040
1539 QISEVOENVLQKFEHRKANDSALOSTESKMLTJNRLQSOEBOEIQIMIKEL----- 1592
1041 KLOEQSDLEQERLAKELQEQSDLE--QERLAKELQEQSDLEQERLAKELQEQSDL 1098
1593 --EEMKRVQALQIERDQKENTKEIVAKMK-ESQEKYQFLQMTAVNETQEKMCIEH- 1648
1099 DLEQERLAKELQEQSDLEQ--ERLAKELQEQSDLEQERL-AKEKLOEQSDLEQ 1154
1649 --LKEQFQTKMLNENIETNIRLTQILHNLEEMSVYKERDRLSVEETLAKVERDQK 1706
1155 RLAKELQEQSDLE-----QERRAKEKLOEQSDLEQERLAKELQEQSDLEQER- 1209
1707 ENLRETTITRLKEQELKIVHMLKEHOETIDKLGIVSEKTNESNMQKXLE---HSND 1763
1210 EKLOEQ-QSDLEQERRAK-----EKLOEQSDLEQER-LAKEKLOEQ--QSDLEQERRAKE 1261
1764 ALKAQDLKIQEELRIAHMLKEQOETIDKLGIVSEKTDKLSNMQKXLENSNAKLQKIQ 1823
1262 KLOEQSDLEQERR-AKEKLOEQSDLEQER---RAKEKLOEQSDLEQERL-AKEKLO 1315
1824 ELKANEHQILTLKQVNETOKVSEMEQLKKQIQDQSLTSLKLEIENLNLAQELHENLE 1883
1316 EQSDLEQERLAKELQEQSDLEQERRAKEKLOEQ--SDLEQERL-----AKEKLOE 1367
1884 MKSVKERNLRRVEETLK-----LERDOL-KESLOETKARDLEIQOELKTARMLSEKH 1937
1368 QSDLEQ--RAKEKLOEQSDLEQERLAKELQEQ--QSDLEQERRAKEKLOEQSDL 1423
1938 ETVDKREKISEKTIQISDIQKQDQKDELQKIQELQKLOQLLRKEDVNMHKKIN 1997
1424 EQERRAKEKLOEQ--QSDLEQEQ--RAKEKLOEQSDLEQER---RAKEKLOEQSDL- 1474
1998 EMEQKIQFEPNYLCKEMDNFQTKLHESLEIRIVAKERDELRIKESLKWEDQFI 2057
1475 EQERLAKEL-----KLOEQORDLEQERRAKEKLOEQSDLEQERR--- 1513
2058 ATLREMTARDNRQHVK--PEKRLSDGQOHLMESLREKSCSRIKELLKRYSEMDDHYECL 2115
1514 -----AKEKLOEQSDLEQERLANEKLOEQORDLEQ--RAKEKLOEQSDL- 1557
2116 NRLSLDLKEIEFHRIMKKLYVLSYVTIKKEQHECINKPEMDFIDEVQKELLIKIQ 2175
1558 -----DLEQEQ-----EKLOEQ--SDLEQERRAKEKLO 1587
2176 HLQDCDVPSELRLDKL-NQNMDLHIEILKD--PSESRFPSTKTEFOQVLSNRKEMT 2231
1598 --EQSDLEQERLAKELQEQORDLEQERLAKELQEQORDLEQORDLEQORDLEQORDLEQ 1645
2232 QFLEWLNTRFDIEKNGIQKENDR-----ICQVN--NFFNNRIIIMNESTEF 2280
1646 DVLAEDLYGRLEIPAIE--LPSENERGYIYPHQSILPQDNRGNSRDSKEISIIEXT--N 1700
2281 ERSATISKWEQDLKS--LKERNEKLFKNYQTLKSLASQAVNPTTQDNKNPHVTSRAT 2338
1701 RESITTNVGRDIIHGHLKEKKD-----GSIKPEQEKDSADIONHTL 1744
2339 Q-LITTEKRELENSLHEAKESAMHKEKLIIMQKLEVTINDIIAKLQAKVHESNKCLEKT 2397
1745 ETVNISDVNDFQISKYEISAIYDDSLIDE-----EEDDEDLDEF 1785
2398 KETIQV--LQDKVALGAKYKEBIEIDLKMLKVLKIDLEKMKNAKEFEKISATKATVYOK 2455
1786 KPIVQYDNFQDEBNIGI--YK-ELEDL-----LEKNENLDLDEGI-----EK 1825
2456 EVIRLRENLRSQQAQDTSVISEHTDPOPSNKP LTCGGSGGIVQNTKALILKSHIRLE 2515

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:30:02 ; Search time 47.178 Seconds
(without alignments)
2914.068 Million cell updates/sec

Title: US-10-045-631B-88

Perfect score: 13329

Sequence: 1 MAEGAVAVCVVRPLNSR.....SQPCWHASSGKDVPCKTQ 2663

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13313.5	99.9	2662	4	US-09-595-684B-31
2	3615.5	27.1	2954	4	US-09-150-867-1
3	1077.5	8.1	1388	4	US-09-572-191-2
4	1077.5	8.1	1388	4	US-09-723-262-2
5	1077.5	8.1	1388	4	US-09-723-219-2
6	1052.5	7.9	3878	4	US-09-914-259-11
7	1040.5	7.8	3248	1	US-08-353-700-1
8	1040.5	7.8	3248	5	PCT-US95-16216-1
9	1037	7.8	2482	1	US-08-328-254-6
10	946	7.1	1231	4	US-09-595-684B-23
11	941.5	7.1	1232	4	US-09-592-054-2
12	924.5	6.9	1234	4	US-09-592-054-8
13	889	6.7	1375	4	US-09-722-139-2
14	889	6.7	1375	4	US-09-721-832-2
15	889	6.7	1375	4	US-09-721-689-2
16	867.5	6.5	963	4	US-09-914-259-20
17	867	6.5	1898	1	US-08-056-200-94
18	867	6.5	1898	2	US-08-800-644-94
19	867	6.5	10182	4	US-09-134-001C-3159
20	866.5	6.5	957	4	US-09-914-259-16
21	863	6.5	956	4	US-09-914-259-17
22	856	6.4	967	4	US-09-914-259-21
23	848.5	6.4	963	4	US-09-914-259-22
24	845.5	6.3	1279	4	US-09-724-517-2
25	845.5	6.3	1279	4	US-09-641-807A-2
26	845.5	6.3	1279	4	US-09-723-096-2
27	844.5	6.3	975	4	US-09-914-259-19

28	844	6.3	1031	4	US-09-914-259-24
29	837	6.3	1027	4	US-09-914-259-27
30	835.5	6.3	1032	4	US-09-914-259-26
31	830	6.2	1066	3	US-09-541-782-8
32	830	6.2	1066	4	US-09-723-820-8
33	830	6.2	1066	4	US-10-270-085-8
34	824	6.2	2310	4	US-09-874-923-120
35	803.5	6.0	935	4	US-09-914-259-25
36	800.5	6.0	3696	4	US-09-134-001C-5080
37	782.5	5.9	1972	4	US-08-875-435B-3
38	779	5.8	1057	4	US-09-428-156B-2
39	778	5.8	1637	4	US-09-718-692-2
40	778	5.8	1637	4	US-09-718-852-2
41	778	5.8	1637	4	US-09-718-815-2
42	776.5	5.8	1056	4	US-09-595-684B-29
43	774.5	5.8	1972	4	US-08-875-435B-4
44	770	5.8	2101	1	US-08-466-390-4
45	770	5.8	2101	1	US-08-470-950-4

ALIGNMENTS

RESULT 1

US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766

; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe

; APPLICANT: Ohashi, Cara

; APPLICANT: Sakowicz, Roman

; APPLICANT: Vaisberg, Eugeni

; APPLICANT: Wood, Kenneth

; APPLICANT: Yu, Ming

; TITLE OF INVENTION: Human kinesins and methods of producing

; FILE OF INVENTION: and purifying human kinesins

; FILE REFERENCE: cytop036

; CURRENT APPLICATION NUMBER: US/09/595,684B

; CURRENT FILING DATE: 2002-06-24

; PRIOR APPLICATION NUMBER: 09/295,612

; PRIOR FILING DATE: 2000-04-20

; NUMBER OF SEQ ID NOS: 105

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 31

; LENGTH: 2662

; TYPE: PRT

; ORGANISM: Human

; US-09-595-684B-31

Query Match

Best Local Similarity 99.9%; Score 13313.5; DB 4; Length 2662;

Matches 2662; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY	1	MAEGAVAVCVVRPLNSR	ESLGTAQVYWKTDNNVIYQV	SGSKSFNDRVPHGNETTK	60
DB	1	MAEGAVAVCVVRPLNSR	ESLGTAQVYWKTDNNVIYQV	SGSKSFNDRVPHGNETTK	60
QY	61	NYEETIAAPIIDSAIQGYNGTIFAYGQTASGKT	YTMGSEDLHGVIPRAIHDFQIKKF	120	
DB	61	NYEETIAAPIIDSAIQGYNGTIFAYGQTASGKT	YTMGSEDLHGVIPRAIHDFQIKKF	120	
QY	121	PREPFLRVSYMEIYNETITDLCTQKMKPLIREDVNRN	VYVADLTTEVVYVTSMAJK	180	
DB	121	PREPFLRVSYMEIYNETITDLCTQKMKPLIREDVNRN	VYVADLTTEVVYVTSMAJK	180	
QY	181	WITKEKSRHVGETKQNRSSRSHTIFRMILBSREKGRPS	NCGSKVYSHLNLVDLAGE	240	
DB	181	WITKEKSRHVGETKQNRSSRSHTIFRMILBSREKGRPS	NCGSKVYSHLNLVDLAGE	240	
QY	241	RAAQGTGAAGVRLKEGNCINRSLFILGQVKKLSDQVGV	GFINRYDSKLTIRLQNSLGNP	300	
DB	241	RAAQGTGAAGVRLKEGNCINRSLFILGQVKKLSDQVGV	GFINRYDSKLTIRLQNSLGNP	300	

QY 301 KTRIICTITPVSPDITLALQFASTAKYKMTPTVYNEVSTDEALLKRYKEIMDLKKOLE 360
 Db 301 KTRIICTITPVSPDITLALQFASTAKYKMTPTVYNEVSTDEALLKRYKEIMDLKKOLE 360
 QY 361 EVSLETRAQAMEKDQALQLEEKDOLLQKVQNEKIENITRMLVUTSSITLQOELKAKKRR 420
 Db 361 EVSLETRAQAMEKDQALQLEEKDOLLQKVQNEKIENITRMLVUTSSITLQOELKAKKRR 420
 QY 421 VTWCLGKINOMKSNYADQFNIPNTITTKTKHLSINLLREIDESCSDVFSNTDLS 480
 Db 421 VTWCLGKINOMKSNYADQFNIPNTITTKTKHLSINLLREIDESCSDVFSNTDLS 480
 QY 481 EIEWNPATKLNQENIRSELNSLRADYDNLVDYEQLFTEKEEMELKKEKNDLDEFAL 540
 Db 481 EIEWNPATKLNQENIRSELNSLRADYDNLVDYEQLFTEKEEMELKKEKNDLDEFAL 540
 QY 541 ERKTKKQDEMOLIHIEISNLKNLVKRVYVQNDLENELSSKVVELLREKEDQIKKQYIDS 600
 Db 541 ERKTKKQDEMOLIHIEISNLKNLVKRVYVQNDLENELSSKVVELLREKEDQIKKQYIDS 600
 QY 601 OKLENIKMDLSYSIESIEDPKOMKOTLPDAETVALDAKRESAFLSENLELKEKMKELAT 660
 Db 601 OKLENIKMDLSYSIESIEDPKOMKOTLPDAETVALDAKRESAFLSENLELKEKMKELAT 660
 QY 661 TVKOMENDIQYQSOLEAKKQVQDLKELQSAFNEITKLTSLIDGKVPKDLCLNLELEG 720
 Db 661 TVKOMENDIQYQSOLEAKKQVQDLKELQSAFNEITKLTSLIDGKVPKDLCLNLELEG 720
 QY 721 KITDLOKELNKEVEERNEALREEVILLSELKSLPSEVERLRKEIQKSGHELHIITSEKDL 780
 Db 721 KITDLOKELNKEVEERNEALREEVILLSELKSLPSEVERLRKEIQKSGHELHIITSEKDL 780
 QY 781 FSEVVKHSRVGCLLEIETKTDLATTOSENKSTDDQFQNFKTLMHDPQKYKWLLEN 840
 Db 781 FSEVVKHSRVGCLLEIETKTDLATTOSENKSTDDQFQNFKTLMHDPQKYKWLLEN 840
 QY 841 ERMNOEIVNLSKEAQFSSIGALKATLSYKTOELQEKTRVQERLNEMEQKLEQENRD 900
 Db 841 ERMNOEIVNLSKEAQFSSIGALKATLSYKTOELQEKTRVQERLNEMEQKLEQENRD 900
 QY 901 SPLQTVREKTLITELKQOTLBEVKTLTQEKDOLLQKQESLQIERDOLAKSDIHDVTNMNI 960
 Db 901 SPLQTVREKTLITELKQOTLBEVKTLTQEKDOLLQKQESLQIERDOLAKSDIHDVTNMNI 960
 QY 961 DTQEQLRNALESIKHOETINTLTKSKI SEEVSNLHMEENTGRTKDEFOQKVGIDKKQD 1020
 Db 961 DTQEQLRNALESIKHOETINTLTKSKI SEEVSNLHMEENTGRTKDEFOQKVGIDKKQD 1019
 QY 1021 LEAKNTQTLTADVKNONEIIEQQRKIFSLIQEKNELOQMLSVIAEKEQKTDLKENIEMT 1080
 Db 1020 LEAKNTQTLTADVKNONEIIEQQRKIFSLIQEKNELOQMLSVIAEKEQKTDLKENIEMT 1079
 QY 1081 IENQEBELRLIGDELKQOEIVAOEKHNAIKKEGELSRDCLRLAEVBEKLEKSOOLEKQ 1140
 Db 1080 IENQEBELRLIGDELKQOEIVAOEKHNAIKKEGELSRDCLRLAEVBEKLEKSOOLEKQ 1139
 QY 1141 QOLANQOEEMSEMOKKINEIENLKNELKNEKELTLEHMETRLELAOKLNEYEVKSITK 1200
 Db 1140 QOLANQOEEMSEMOKKINEIENLKNELKNEKELTLEHMETRLELAOKLNEYEVKSITK 1199
 QY 1201 ERKVLKELQKSFETERDHLRGYIREIBATGLQTKGELKIAHILKHQETIDELRSVSE 1260
 Db 1200 ERKVLKELQKSFETERDHLRGYIREIBATGLQTKGELKIAHILKHQETIDELRSVSE 1259
 QY 1261 KTAQIINTQDLESHTKLOBEIPVLEHEQHLNPKVKSSETQETMNELELLTQSTTKDS 1320
 Db 1260 KTAQIINTQDLESHTKLOBEIPVLEHEQHLNPKVKSSETQETMNELELLTQSTTKDS 1319
 QY 1321 TTILARIEMERLRLNEKQESQEEIKSLTKERDNLTKIKEALEVKHQDLKSHIRETLAKIQ 1380
 Db 1320 TTILARIEMERLRLNEKQESQEEIKSLTKERDNLTKIKEALEVKHQDLKSHIRETLAKIQ 1379
 QY 1381 ESQSKQPCSLNMKSKONETTKIVSEMEQFPKDSALLRIEIMLGLSKRLOESHEDEKMSV 1440

Db 1380 ESQSKQPCSLNMKSKONETTKIVSEMEQFPKDSALLRIEIMLGLSKRLOESHEDEKMSV 1439
 QY 1441 AKEKDDIQRLOQEVLOQESDQKLENIKIIVAKHLETEBELKVAHCLKEQBETINELRVNL 1500
 Db 1440 AKEKDDIQRLOQEVLOQESDQKLENIKIIVAKHLETEBELKVAHCLKEQBETINELRVNL 1499
 QY 1501 SEKETEISTIQKOLEANDKLONKI OELIYKEKEOLNIQIISFVQBNVNLQKQFHEKRAK 1560
 Db 1500 SEKETEISTIQKOLEANDKLONKI OELIYKEKEOLNIQIISFVQBNVNLQKQFHEKRAK 1559
 QY 1561 DSALQSTESKMLLETNRLQESQEEIQIMIKEBEMKRVQEAQOIERDOLKENTKEIIVAKM 1620
 Db 1560 DSALQSTESKMLLETNRLQESQEEIQIMIKEBEMKRVQEAQOIERDOLKENTKEIIVAKM 1619
 QY 1621 KESQEKYQFLKMTAVNETQKMCIEIHLKEQFQTKLNLNIENTENIRLTQILHENLEE 1680
 Db 1620 KESQEKYQFLKMTAVNETQKMCIEIHLKEQFQTKLNLNIENTENIRLTQILHENLEE 1679
 QY 1681 MRSVTKERDDILRSVEETLKVERDOLKENLRSETITRDLEKQEBELKIVHMLKHEQETIDKL 1740
 Db 1680 MRSVTKERDDILRSVEETLKVERDOLKENLRSETITRDLEKQEBELKIVHMLKHEQETIDKL 1739
 QY 1741 RGIYSEKTNELSNMOKDLEHNDALKAQDLKI QEBELRIAHMLKHEQETIDKL RGIYSEK 1800
 Db 1740 RGIYSEKTNELSNMOKDLEHNDALKAQDLKI QEBELRIAHMLKHEQETIDKL RGIYSEK 1799
 QY 1801 TDKLSNMOKDLEHNDALKAQDLKI QEBELRIAHMLKHEQETIDKL RGIYSEK 1860
 Db 1800 TDKLSNMOKDLEHNDALKAQDLKI QEBELRIAHMLKHEQETIDKL RGIYSEK 1859
 QY 1861 LTLKLEIENLNLQELHENLEEMKSVMKERDNLRSVEETLKBRDOLKESQETIKAROL 1920
 Db 1860 LTLKLEIENLNLQELHENLEEMKSVMKERDNLRSVEETLKBRDOLKESQETIKAROL 1919
 QY 1921 EQQELKLTARMLSKHETVDKLREKISEKTIQISDIQKDLKSKDELQKQI OELQKKE 1980
 Db 1920 EQQELKLTARMLSKHETVDKLREKISEKTIQISDIQKDLKSKDELQKQI OELQKKE 1979
 QY 1981 QLLRVKEDVNMGHKKINEMEQKQFEPNYLCKCEMDNPLTKLHESLEEIRIVAKED 2040
 Db 1980 QLLRVKEDVNMGHKKINEMEQKQFEPNYLCKCEMDNPLTKLHESLEEIRIVAKED 2039
 QY 2041 ELRRIKESLAKWERDQFIATLRMIARDRONHVKPEKRLSDGQOQHLMESLEKCSRIKE 2100
 Db 2040 ELRRIKESLAKWERDQFIATLRMIARDRONHVKPEKRLSDGQOQHLMESLEKCSRIKE 2099
 QY 2101 LLKRYSEMDDHYECLNRLSLDLEKETEFRIMKKLYVLSYVTKI KEEQHECINKPEMDF 2160
 Db 2100 LLKRYSEMDDHYECLNRLSLDLEKETEFRIMKKLYVLSYVTKI KEEQHECINKPEMDF 2159
 QY 2161 IDEVEKQKELLIKIQHLOQDCVPSRELDRDLKLNQMDLHIEILKDFSESEPPSITKTEF 2220
 Db 2160 IDEVEKQKELLIKIQHLOQDCVPSRELDRDLKLNQMDLHIEILKDFSESEPPSITKTEF 2219
 QY 2221 QQVLSNRKEMTQPLEEWLNTFRDIEKLNKGIQKENDRIQVNNFFNNRIIAINNESTEFE 2280
 Db 2220 QQVLSNRKEMTQPLEEWLNTFRDIEKLNKGIQKENDRIQVNNFFNNRIIAINNESTEFE 2279
 QY 2281 ERGATISKEWEOQLSKKEKNEKLFKNYQTLTKTSLASGAQVNPPTQDNKNPHVTSRATOL 2340
 Db 2280 ERGATISKEWEOQLSKKEKNEKLFKNYQTLTKTSLASGAQVNPPTQDNKNPHVTSRATOL 2339
 QY 2341 TTEKIRELENSLHAEKESAMHESKIIKMQKELEVNDIIAKLOAKVHESNKCLEKTKET 2400
 Db 2340 TTEKIRELENSLHAEKESAMHESKIIKMQKELEVNDIIAKLOAKVHESNKCLEKTKET 2399
 QY 2401 IOVLQDKVALGAPYKEEIEDLAKMLVKIDLEKMKNAKEFEKESATKATVEYQKEVIRL 2460
 Db 2400 IOVLQDKVALGAPYKEEIEDLAKMLVKIDLEKMKNAKEFEKESATKATVEYQKEVIRL 2459
 QY 2461 LRNLRRSQOAOQTSVISEHTDPQSNKPLTCGGGSGIVONTKALILKSHIRLEKISK 2520


```
Db 2460 LRENLRSSQAQDTSVISEHTDPQSNKPLTCGGSGIVQNTKALILKSHIRLEIKS 2519
Qy 2521 LKQONEOLIKQKNELSNNOHLNNEVKTWKERTLKRRAHQVTCENSPKPKVTGTASK 2580
Db 2520 LKQONEOLIKQKNELSNNOHLNNEVKTWKERTLKRRAHQVTCENSPKPKVTGTASK 2579
Qy 2581 KQITPQCKERNLQDPVPKESPKSPFDSRSKSLPSPPHVPVRYFDNSSLGLCPEVQNAE 2640
Db 2580 KQITPQCKERNLQDPVPKESPKSPFDSRSKSLPSPPHVPVRYFDNSSLGLCPEVQNAE 2639
Qy 2641 SVDSQCPWHASSGKDVPECKTQ 2663
Db 2640 SVDSQCPWHASSGKDVPECKTQ 2662

RESULT 2
US-09-150-867-1
; Sequence 1, Application US/09150867
; Patent No. 6645748
; GENERAL INFORMATION:
; APPLICANT: Wood, Kenneth W.
; APPLICANT: Sakowicz, Roman
; APPLICANT: Goldstein, Lawrence S.B.
; APPLICANT: Cleveland, Don W.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Plus End-Directed Microtubule Motor Required for
; FILE REFERENCE: 18557C-000110US
; CURRENT APPLICATION NUMBER: US/09/150,867
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: US 60/058,645
; EARLIER FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2954
; TYPE: PRT
; ORGANISM: Xenopus sp.
; FEATURE:
; OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)
; OTHER INFORMATION: member of the kinesin superfamily of microtubule
; OTHER INFORMATION: motor proteins
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1)..(472)
; OTHER INFORMATION: kinesin like motor domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (473)..(2752)
; OTHER INFORMATION: rod domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (2753)..(2954)
; OTHER INFORMATION: tail domain
US-09-150-867-1

Query Match 27.1%; Score 3615.5; DB 4; Length 2954;
Best Local Similarity 32.0%; Pred. No. 5 4e-162;
Matches 99%; Conservative 572; Mismatches 932; Indels 615; Gaps 86;

Qy 1 MAEGAVAVCVVRPLNSREESLGETAQQVYWKTDNNVYQVDSKSFNDFVPHGNETTK 60
Db 1 MSEGAVKCVVRPLIQREQ-GDQANLQWAGNNTISQVDGTSKSFNDFVFNHSHTS 58
Qy 61 NYVERIAAIPIDSAQGYNGTTFAYQGTASGKTYTMGSEDHGLGVIPRAIHDFOKIKKF 120
Db 59 QIQEIAVPIIISALQGYNGTTFAYQGTSGKTYTMGTPNSIGIIPQAIQEVFKIIEI 118
Qy 121 POREFLLRVSYMEIYNETITDLLCGTKQMKPLIREDVNRNVYVADLTVEVYVTSMAK 180
Db 119 FNRFLLRVSYMEIYNETIVKDLCCDDRRKKPLIREDVNRNVYVADLTVEVYVPHVIQ 178
Qy 181 WITGKESRHYGETKMNORSSRSHTIFRMILESREKGEPS---NCEGSKVSHLNLVDLA 237
```

```
Db 179 WIKKGEKNRYGHTKMDHSRSHTIFRMIVESDRNDPTNSENCDGAMVSHLNLVDLA 238
Qy 238 GSERAQGTGAAGVRLKEGCNINRSLFILGVIKKSDGVGGFINYRDSKLTRIQNSLG 297
Db 239 GSERASQTGAAGVRLKEGCNINRSLFILGVIKKSDGVGGFINYRDSKLTRIQNSLG 298
Qy 298 GNPKTRIICTITVPSDETLTALQFASHTAKYMKNTPYNEVSTDEALLKRYRKEIMDKK 357
Db 299 GNAKTVIICTITVPSDETLTALQFASHTAKYMKNTPYNEVSTDEALLKRYRKEIMDKK 358
Qy 358 QLE--EVSLETRAQAMEKQOLAQLLEKOLLQVQNEKIENLJEMLVTSLSLTQQLKA 415
Db 359 QLENLESSETKAQAMAKEHTQLAEIKOLHREDRIWHLTNIVVASSQES-QQQRV 417
Qy 416 KRKRVTWCLGKINKMKNYADQFN---IPTNITTKTKLSINLREIDSVCSSESDV 471
Db 418 KRKRVTWAPGKIQNSLHASGVDFDMLSLPGNFSKAKFSDMPSPFPEIDDSVCTEFS 477
Qy 472 FSNLTDTLS---EIEWNPATKLNQE----- 494
Db 478 FDDALSMDNSGIDAENLASKVTHREKTSLSHSMIDFGQISDSVQFHDSSKENQLQYLP 537
Qy 495 -----NIESE 499
Db 538 KDSGDMAECRKASFEKEITSLOQLOSKREKKELVQSPFELKAELEEQLSVAKNLEMV 597
Qy 500 LNS----- 502
Db 598 TNSRHSINAQVQDVEKVRKEMSVLGDGYNASNSLDQDSSVDGKRLSSSHDECIEH 657
Qy 503 -----LRA 505
Db 658 RKMLEOKIVDLEEFINLNKKGSKSSEQDFMESIQLCEALMAEKANALEFALMRD 717
Qy 506 DYDNVLVDYEQLRTEKEEMELKKNLDLDEFEALERTKTKDOEMOLIHESLNKLVKH 565
Db 718 NFDNIIENETLKRRIADLERSKENQETNEFEILEKETQKEHEAQLIHEISGLKLVEN 777
Qy 566 REVYNODLENLSKVELLREKEDQIKLQEVIDSQLENINIKMDLSVSL-----ESIEDPK 621
Db 778 AEMYNQNLSEDLTKLKEQEIQLAELRKADNLQKVRNFDLSVSGDSEKLCIEIF 837
Qy 622 QMKOTLFDATVVALDAKRESAFLRSENLELKEKMLATYKQMDNDIQLYSQLEAK-- 679
Db 838 QLKQSLSDAEAVTRDAQKESFLRSENLELKEKMDTNSWYNQKKAASLFEKQLETEKS 897
Qy 680 --KMQVDLEKELQSAFNEITKLTSLIDGKVPKDLLCNLELGKITDLOKELNKVEENE 737
Db 898 NYKKMEADLQKELQSAFNEINVINGLAGKVPFDLLSRVELEKKVSEFSKQLEKALEKN 957
Qy 738 ALREEVILLSELKSLPSEVERLRKIQDKSEELHIITSEKDLFSEVHVHESRVQGLLEE 797
Db 958 ALENEVTCLSYKFLPNEVECLKNQISKASEEIMLLKQSEHSASIIISKQEIIMQSQSEQ 1017
Qy 798 ICKTKDGLATTOSNYKSTDOEFQNKTLHMDPEQKVMVLENERMNOQIVNLKSAQKF 857
Db 1018 ILQLDEVHTHTOSKVOQTEEQYLEMKKMDHDLFEKY-----IRNKSEADLLREMN 1070
Qy 858 DSSLGALTKELSYKQELQKTEVQERLNEQLEKEQLEN-RDSPLOTVREKTLITE 915
Db 1071 KGTMESVEVKIADTKHELEETIRDEQLLHEKKYFFQAMOTIFPTPIPSDSLPSPKLVGG 1130
Qy 916 KLQQTLE--EVKTLTQEKDDLKQLESQIERDQKLSDIHDTVNNNIDTQELRNALES 973
Db 1131 NSQDPIEDVHNLTALATERNINMVCLTERNSLKEQV-----IDLNTQLQ-SLQQA 1182
Qy 974 KQQTINTLKSISE-EVSRNLHMEENTGETKDFQKMGVIDKKQ--DLK-AKNTQTL 1029
Db 1183 STEKSDLPKQKQDLKEGEVKKLLEMLKGLHTD---SLSIEKLQLENLEVTETKQTL 1238
Qy 1030 TADVKNDEIIEQQRKIFSLIQKELQKLESVIAKEQLKTDLENEMTLENQELRL 1089
```

1239 QEBMKNIT-----ERNELQTNFEDLKAEDSLKQDLSNENIOSIETQDELRA 1286
 1090 LGDELKQOEIVAEQKHAIK-KEGELSRCDRLAEVEEKLKEKSOQLQKQOQLLNQVE 1148
 1287 AQEELREKQQLVDSFRQOOLDCSVGISPNHDVANGQKVSGLGVNSLOSEM-----LRG 1341
 1149 EMSQMKK-----INEIENLNKELNKLTLHEMETERLELAQKLNENYEVKSTIKERKV 1204
 1342 ERDELQTSKRALVSELELLRAHVS-----VGENLEITKLNGLKEKEILGKSESEV 1394
 1205 LKELOKSPETERDHLRGV-----IRETEATGLQTKELKTAHILKHEQHTID 1252
 1395 LKSMLENLKEDNNKLKEQAEYSSKENQFSLSEVFSQKLVDIEVLKALKAABERLE 1454
 1253 ELRSVSEKTAQIINTQDLE-KSHTKLOEIPVLHHEQELLPNVKVSETQETWNELELL 1311
 1455 IKRDYFE-LVQTANTNLVEGKLTPLQAD-----HEED-----SIDRRSEEM-EIKVL 1501
 1312 TEQSTTKDSTTLARIEMERLRNKEKFBQSEBEIKSLTKERDNLKTIKALEVKGHDQKHEH 1371
 1502 GEK-LERNOYLERLQEBKELSNKLETLQEMETSVLLKDDLOQLKESLSLSENIILKEN 1560
 1372 IRETL-----AKIOESQKQOSLANKEKDNETTKIVSEMEQPKPKDSALLRTEIEML 1424
 1561 IDTTLKHSDDTQAOLQKTQOQLQALAKNLAIAASONCPITQKE-----TSADCVHPLLEKIL 1617
 1425 GLSKRLOESDBMKSVAKEKDDLOQLQVLOSEDOL-----KENIKETIVAKHLETEE 1477
 1618 LLTEELHOKTNEQKLLHEKNELEQAQVELKCEVEHLMKSMLESKSSLESLOEKHDTEQ 1677
 1478 ELKVAHCCLEKQEBETINELRNVLSKE-----TEISTIOKQLEBAINDKLNKIQEIYEKBE 1533
 1678 QL-----LALKQMQVVTQEKKELOQTHELTAEDVHLKENIE-LGLNFKNEAQQTKTKEQ 1732
 1534 QL-NIKQISQVEQNNV-ELKQFKEHRKAKDSALQSI---BSKMLELNRLOESDEEIQI 1587
 1733 CLINENKLEOSQHLQCEIEELMKSLOKDSALETLKESEOKVINL-----NQEMEM 1785
 1588 MIKEEEMKRVQEAQIQRDOLKENTKEIVAKMKESQE---KEYQFLKMTAVNETQKMC 1645
 1786 VMLEEELKNSQRTVIARDQLODDLRESVEMS IETQDDLKAQEAQKQKQVQELTSQ 1845
 1646 IEHLKEQFETQKLNLIETENIRUTQILHENLEEMRSVTKERDRLRSVEETLKVERQOL 1705
 1846 ISVLQEKISL-----LEN-----QMLY-NVATVKETLSRDDLQNSKOHLFSEIETL 1891
 1706 KENLETTITRDLKEQELKIVHHLKHEQHTID---KLGRIVSEKTNESNMOKDLEHNS 1762
 1892 SLSLKEK-EFALQEAQKDK-----ADAAKTIIDITEKISNIEEQQLQOAINLKEITL-YER 1944
 1763 DALKAQDLKIQELRLAHMHLKEQOETIDKLGRIVSEKTDKLSNMOKOLENSNAKLOEKI 1822
 1945 ESL-----TQCKQLALNTEHLRETLKSLDALGKMEQERDEAANKVIALTERKMSLSLEQI 2000
 1823 QE-----LKANEHOLITLKQVNETQKVSSEMOLKQIKQOQSLTSLK-----HIE 1869
 2001 NENVITLKEGEKETFYLRPSKQOSSQMBELRESLTKDLOLEAEKEITSEATNEIK 2060
 1870 NL-----NLAQELHENLEEMKSVMKERDNLRRVEETLKLERDOLKESLOETKARDLEIQ 1924
 2061 NLTAKISSLEETLOQASILNEAVSERENLRHSKQOLVSELQSLTUL---KSRDHAFQ 2117
 1925 ELKTAARMLSKHEKTVDLKREKISIKTIQISIQKDLK---SKBELQKKIQELQ----- 1976
 2118 -----SKREK---DEAVNKIASLAEBEIKULTKEMDFRDSKESLSQSSHLSBELCT 2166
 1977 -KKELOLLR-VKEDVNMGS-HKKINEMEQLKQEPNLYLKCENMDNFQTKKLHESLEIR 2033
 2167 YKTELQMLKQKEDINNKLAEKVEYDEL-----LQHLSSKLEQDQIQ 2210
 2034 IVAKERDELRRIKESLQME-RDOFIATLEMIARDRQNHQVKEPKLLSDGQOHLMESLR 2092
 2211 MELR-NEKLNRYELCEKMDIMEKISVLRLM-----QNEPQOE-EDDVAERMDILE 2259

2093 EKCSRIKELLKRVSEM-DDHYECLNRLSLDLKEKEIEHR----- 2130
 2260 SRNQEIQELMEKISAVTSEQHTLSSLSSELOKETEAHKHOMLNIMKESLSSTLSRSFGSL 2319
 2131 -----IMKKLVLSYVTKIKESQHCINKFEMDFIDEVEKQKELLIKIQLH 2177
 2320 QTEHVKLNTQLOTLNLFKVV--YRTAAVEDHSLIKDYEKDLAAEQKRDHDLQLOQL 2377
 2178 QDQ-----CDVPSRELRLKLNQNMNL-----HTERILKOFSESE--FPSIKTFEQVLS 2225
 2378 EQHGRKWSDSASELKECEIEFLNELLFKKANIISQVDDDFSEVQVFLNQVSGTLOEEL 2437
 2226 NRKEMTOFLBWLNRFDIEKLNKGIQKENDRICOVANFFNRRNRIIATINESTEPERSAT 2285
 2438 HKGFQMWLEEFGLDHLVDKKLSEGMOQENRRRIASTTQLLTKRLKAVVQSKI---QREIT 2494
 2286 I-SKEWQDLKSLKEKNEKLFKNYQTLKTSLSAQVN-----PTQDNKNPHVTSRA 2337
 2495 VYLNQFEAKLOEKEQNKELMRMEHGPSVMEENARLLGILKTVQDE-----SKK 2548
 2338 TOLTEKIRELENSLHBAKESAMKESKIKMOQKELEVTNDIIAKL---QAKVHESNKCL 2394
 2549 LQ-----SRIKMLENELNLVKDDAMHKGKVAILODKL-LSRNAEAEALNAMQVLTCKKQDNL 2604
 2395 EKTETITQVLQDVALGAKPYKEITEDLKMKLVKIDLEKMKNAKEFEKEISATKATVEYQ 2454
 2605 QAAMKEITENLOKMAKAVPYKEIDNLKTVKVIKLEMEKIKYKATDOEIAVLKSLCLBCK 2664
 2455 KEVIRLLRENRRSQQAODTSVISEHTDPQPSNKPFLTCGGSGGIVQNTKALILKSEHRL 2514
 2665 EGBURLKEELRRAQANDTTVCVPKDYQKASTPVTTCGGSGGIVQSTAMLVLQSEKAA 2724
 2515 EKEISKLKQONEQL-----IKQKNELLSNNOHLSNEVKTWERTILKRAHKQV 2562
 2725 ERELSHYKKYVHLSRTMSSSEDRKTKAKSDAHSHTGSSHRGSPHKTETVYR---HGPV 2781
 2563 TCE-----NSPK-----SPKVTCT-----ASK 2579
 2782 TPERSEMPSLHLGSPKSESSTKRVVSPNRSEIYQVMSPOKTGMHKLHILSPSKVGLHK 2841
 2580 KKOITP-----SCKERNLODPVPKESPKSCFF 2607
 2842 KRAISPNESEMTQHVISPGKTGLHKNLTESTLFDNLSSPCQKQKQVQENL--NSPKGLF 2899
 2608 DSRKSLSPHPVRVFDNSSLGLCEPVONAGAEVDSQGPWHASSGKDVPECKT 2662
 2900 DVKSKSMFY-CPSQPFDSNKLGLDFSELNTAESNDKSAENMMWYEAKETAPECKT 2953

RESULT 3

US-09-572-191-2
 ; Sequence 2, Application US/09572191
 ; Patent No. 6355466
 ; GENERAL INFORMATION:
 ; APPLICANT: Beraud, Christophe
 ; APPLICANT: Sakowicz, Roman
 ; APPLICANT: Wood, Kenneth
 ; TITLE OF INVENTION: No. 6355466el motor proteins and methods for
 ; TITLE OF INVENTION: their use
 ; FILE REFERENCE: 1017
 ; CURRENT APPLICATION NUMBER: US/09/572,191
 ; CURRENT FILING DATE: 2000-05-17
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 1388
 ; TYPE: PRT
 ; ORGANISM: Human
 US-09-572-191-2
 Query Match 8.1%; Score 1077.5; DB 4; Length 1388;
 Best Local Similarity 26.5%; Pred. No. 4.8e-43;

Matches 475; Conservative 280; Mismatches 521; Indels 519; Gaps 79;

```

QY 3 EGAACVAVRVRPLNSRESL-GETAQVYKTDNNVLYQDGS-----KSNFDP 50
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 23 EGAIAKVFVRIRPAERSGGADGE-----QNLCLSVLSSTSLRLHSPKPTFTFD 73
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 51 RVPHGNETTKNVYEEIAAPIDSAIOGYNGTIFAYGQTASGKTYTWMG-----SDHL 103
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 74 HVADVDTTQSSVATVAKSIVSCMSGYNGTIFAYGQTGSGKFTWMPGSEDSFNHLR 133
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 104 GVPIRAIHDF-----QIKKFPDRPFLRLVSVMEIYNETITDLCTQKMKPLIREDV 158
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 134 GVTPRSFYLFLSDIREKEKAGAGKFLCKSCSFTEIYNEQIYDILL--DSASAGLYLREHI 191
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 159 NRNVYVADLTVEVVYVSEMALKWITKCEKSRHYGETKMNQSRSHITFERMILESREKGE 218
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 192 KGVFVVGAVEQVVTSAEAYQVLSGGWRNRVASTMNRSSRHAFVTTITTESMEK-- 249
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 219 PSNCEGSKVYSHLNLVDLAGSRAAQTGAAGVRLKEGCNINRSLFIIGQVKKLSGQVG 278
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 250 -SNEIWNRTSLNLVDLAGSRQKQTHAEGMLKEAGNINRSLSCLGQVITALVD--VG 306
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 279 G-----FINTYRDSKLTILQNSLGNPKTRIICTITPVS--FDETALQAPASTAKYMN 332
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 307 NGQRHVYCYRDSKLTFLRLSDLAGNAKTAIANVHPGSRGCFGETLSTLNPQRAKLKKN 366
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 333 PYVNE-----VSTDEALLKRYKEIMDLKK-OLEEVSLETR-----AQAMEKDQALQL-- 380
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 367 AVVNEDTQGNVQIQAEVVRKKEQLAEAGQTPPESFLTRDKKTKNYMEYFOEAMLFFK 426
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 381 ----EKKDLQKQVNEKIEMLVTSSTLTQBELKAKRKRVRVWCLGKINKMNSNY 436
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 427 KSEQENKSLIEKV-----TQLEDLTLLKKEFI-----QSNKMIVKPR 463
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 437 ADFQNPNTWITTKHKLSI-NLRLDEDSVCSSEDSFNSVTLDTLSE-IENNPATKLLNOE 494
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 464 EDQI-----IRLEKLHRESGGFLPEEQDRLLS--LENEIQLRBEQIEHHPVAKYAME 516
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 495 NIESEINSLRADVNLVDYEQRLRTEKEMELKKEKNDLDEFEALERTKQDOEQMLH 554
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 517 N-----HSUR-----EENRRLLE-----PVKRAQEM-- 539
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 555 EISNLKMLVHRVYNQDLENELSKVELLRKEDQIKLQEIYDSQKLENIKMDLSYSL 614
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 540 ----DAQTIAKLEKAF-----SEISG-----MEKSD--KNOQGF----- 567
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 615 ESIEDPKQMKTFLDFAETVALDAKRESAFURSENLEIKERMK-ELATTKYQOMENDIQLYQ 673
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 568 ----SPAKQKEPCLFANT-----EKLKAQLLIQIOTELNNSKQEYE 603
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 674 SLEAKKQKQVLDLEKELQSAFNEITKLTSLDGVKPKDLLCNLESGKITDLOK-ELNKE 732
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 604 EFKELTRKQLELESLSQ-----LQKANLLE 631
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 733 --VEENEALREEVILLSELKSLPSEVERLRKEIQDKSEELHIITSEKDKLFSVHVHKEGR 790
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 632 NLLEATKACKRQ-----EVSQLNK-----IHAETLKIITP-----IKAYQLHSR 671
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 791 VQGLLEIEIKTKDDLTATTSQYKSTQBFQNFKTLHMDPEQKYKVMLEE--NERMNQEI 847
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 672 -----PVPKLPGEMSGFSLYT-----QNSSILDND-----ILNEFPVPPMNEQAF 712
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 848 VNLSKQAKFDSLSIGALKTSELYKTOELOKTRVQERLNEBQLEKQLENRDSPLQTVS 907
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 713 EALSELRTVQOMSAQAKL-----DEEHNKLUQOHVDKLEHHSTQMQLFS----- 762
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 908 REKTLTEKLOOTLEKVTLTQEKDQLQBSLQIERDQLKSDIHD--TVANNIDTQSQ 965
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 763 -----SERIDWKQOBELLSQLNVLEKQLQET-QTKNDFLAKSEVHDLRVLHLSADKE-- 813
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 966 LRNALESILKQOETINTLKSKEEVS-RNLNMEENTGTKEBFQOMVIGIDKKQDLEAK 1024
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 814 ----LSSVKLEYSSFKTQKEKFNKLSRHMV-----QLQDNLRLNEKLESK 860
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 1025 NTQFTLADVQDNELIIBQORKIFSLIOEKNELOQMLSVIAKEQKQKTDLDKENIEMTINQ 1084
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 861 ---ACLOQSDYN--LOEIMKF-----EIDQLSRNLQNFKKENETLKSDLNLMEL-LEAE 909
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1085 EELRLGLDELKQOEIVAOEKHAIKKEGELSRTCDRLAEVEEKLKEKSQOLQEQOOLL 1144
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 910 KE--RNNKLSLQFE--EDKNSK-----ELKVLAVRQKQKETA 947
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1145 NVQEMSEMOKKTINIEINLKNELKELTLEHETERLELAQKLNENYEVKSKTKERKV 1204
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 948 KCEQOMAKVQK-----LE-----ESSLATEKV 969
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1205 LKELQKSPETERDLRGYREIATEGLQTKEEELKIAHILKHEQETIDELRRSVSEKTAQ 1264
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 970 ISSLEKSRSDKKVADLMNQIQ-----ELRSSVCKET-E 1003
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1265 IINT--QDLEKSHTKL-----CEEIPVLHEEQELLPNVKVSETOTMNELELLTQST 1316
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1004 TIDTLQELKDINCKYNSALVDRESRVLKQEV--DIIDLKETL-RLILSE-- 1054
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1317 TKDSTTLARIEMERLRINEKFOSEBEIKSLTKERDNLTKIKALEVYKHDQLKEHIRETL 1376
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1055 -----DIERDMLCEDLAHATEQLNMLTE-----ASKHSGLL-QSAQEEL 1093
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1377 AK-----TORSOSQOEOSIANKKEKDNETTKIVSEMQFPKDSALLRIEIMLGLSKRLOE 1432
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1094 TKKEALIOLOHK-----LNQKKE-----EVEQKQNE-----YNFKMRQ 1127
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1433 SHDEMKSVAKEKDDLQRLQEVLSQESDQLKENTKEIVAKHLET--EELKVHACHLKEQEE 1491
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1128 LEHVMDSAED-----PQSPKTPPHQTHLAKLLETQOEIEDGRASKTSLEH 1175
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1492 TINELRVNISEKETETISTIQKLEAINDKLNQIKQIYEKEEQNLKQISEQVENNELK 1551
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1176 LVTKLNEDREVKNAEILRMKEQLREM-ENLRLESQOLIEKNWLL-----QQQLDDIK 1226
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1552 QKKEHRKAKDSALOSTIESKMLE-LTNRLQESQEEIQIMIKEKEEMKRVQEAL----- 1502
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1227 ROKENSQDHPNQOLKQESQESIKERLAKS-KIVEEMLKMKADLEBEOVALYNKEMECL 1285
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1603 ----QIERDQLKENTKEIVAKMKESQEKYQFLKMTAVNETOEKMC-EIEHLKEQFETQK 1657
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1286 RMTDVEVETQTLSE-----KAFQEKQLBSKLEEMVEEERTSQEEMMLRQ----- 1332
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1658 LNLNETIENIRLT--QILHENLEEMRSVYTKERDDLRSEVEETLKVVERDQKENLR 1710
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1333 --VECLAENGKLVGHQNLHOKIQYVVRLLKENVRLEAETEKLEAENVFLKEKR 1385
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 4

```

US-09-723-262-2
; Sequence 2, Application US/09723262
; Patent No. 6379912
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6379912el motor proteins and methods for
; FILE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/723,262
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
US-09-723-262-2

```

Query Match 8.1%; Score 1077.5; DB 4; Length 1388;
Best Local Similarity 26.5%; Pred. No. 4.8e-43;
Matches 475; Conservative 280; Mismatches 521; Indels 519; Gaps 79;

Qy 3 BEGAVAVCVRPVPLNSREESI-GETAQYVWTKDNNVIYQVDS-----KSFNFD 50
Db 23 EGDAIKVFPVIRPPAERSGSADGE-----QNLCLSVLSSTSLRLHSNPEPKTFTFD 73
Qy 51 RVPHGNETTKNVEIEAIPIDSATIQYNGTIFAYGQTASGKTYTMMG-----SEDLH 103
Db 74 HVADVTITQESVFATVAKSIVESCMGNGTIFAYGQSGKTYTMMGSPSDFNSHLR 133
Qy 104 GVIPRAIHDF-----QXIKFPDPREFLLRVSYMYNETITDLCGQKMKPLIIRDV 158
Db 134 GVIPSPPEYLSLIDREKAGAGKFLCKGKSFIEYNEQIYDVL--DSASAGLYLRHI 191
Qy 159 NRVVYVADLTREVYVTSMAWKWTKGSKSHYGETKQNRSSHTIFRMLISREKGE 218
Db 192 KGVVFGVAVGQVVTSAEAYQVLSGGWRNRVASTSMNRSSRSHTVITIESMEK-- 249
Qy 219 PSNCGSVKVLHNLVDLAGSRAAQTGAAGVRLKEGNCINRSLFPGVVKKLSDGQVG 278
Db 250 -SNEIVNRTSLNLVDLAGSERQKDTABGNLKEAGNINRSLSCIGQVITALVD--VG 306
Qy 279 G-----FINYRDSKLTIRLQNSLGNPKTRIICTITPVS--FDETALQFASKYMKMT 332
Db 307 NGKQHVVCYRDSKLTFLRLDSLGNNAKTAIAIVHPGSRFCGETSLTNFAQAKLIRNK 366
Qy 333 PYVNE-----YSTDAALLKRYKEIMDLKK-OLEEVSLETR-----QAAMEKDOLAQLL-- 380
Db 367 AVVNEDTQGNVSQQAQVRLKEQAEALASGQTPPEFLTRDKKNTNMEYFQEAFLPFK 426
Qy 381 ----BEKLLQKQVQNEKTENLRMLVTSSLTLOELKAKRRRVTCWLGKINKMKNSY 436
Db 427 KSEQKKSLEKV-----TOLEDLTLLKKEKFI-----QSNKMLVKFR 463
Qy 437 AQDFNIPNNTTKTKHLSI-NILREIDSVCSSESVFNTLTLSE-IWNPATKLLNQ 494
Db 464 EDQI-----IRLEKLHESRGGLPBEQDRLLE--LRNEIOTLREQIEHHPVAKYAME 516
Qy 495 NIESELNISRADYDNLVDYEQLRTEKEBEMELKAKENDLDEFEALERTKDKQDMOLH 554
Db 517 N-----HSLR-----EENRRLLE-----PVKRAQEM----- 539
Qy 555 EISNLKNIWKHREVYNQDLENELSKVLELLREKEQIKLQYIDSQKLENIMKLSYSL 614
Db 540 ---DAQTTAKLEAF-----SELG-----MEKSD--KNQGF----- 567
Qy 615 ESIEDPKOMKOTFLDAETVALDAKRESAFIRSENLELKEKMK-ELATTYKQENDIQLYQ 673
Db 568 ---SPKAQKEPCLFANT-----EKLKAQLLIQIOTELNNSQOYE 603
Qy 674 SQLEAKKQVQDLEKELQSAFNEITKLTSLIDGKVPKLLCNLELGGKITDLOK-ELNKE 732
Db 604 EFKELTRKQLESELSQS-----LOKANLNL 631
Qy 733 --VEENHARREEVILLSELKSLPSEVERLRKEIQDKSEELHIITSEKDLFSEVHKEGR 790
Db 632 NLEATKACKRQ-----EVSQNLK--IHAETLKIITP-----TKAYQLHSR 671
Qy 791 VOGLLEIEIKTKDDIATTONSYKSTQDFQNFKTLHMDPEQYKQVLEB---NERMNOI 847
Db 672 -----PVPKLSPEMSPFGSLYT-----QNSILDND-----ILNEFVPPPMNEQAF 712
Qy 848 VNLKSAQKFDLSLKGKLTSELYKTOELQKTRVQERLNEBQLEQLNDRSPLOTVE 907
Db 713 EATSEELRTVOGMSALQAKL-----DEEHNKLLQOHVDKLEHSTQMGELFS----- 762
Qy 908 REKTLITEIKQOOLEEVKTLITQSKDLKQLOESLQIERDQLSDIHD--TVNMNIDTQEQ 965
Db 763 -----SERIDWIKQOEELLSQLNVLKQLOET-QTKNDPLKSEVHDLRVLHSAKDE-- 813
Qy 966 LRNALSLKQHQETINTLSKISEEVS-RNLHMEENTGETKDFQOQKMGVICKQDLEAK 1024

Db 814 ----LSSVKLYESSFKTNOEKFNKLSEHMHV-----QLQDNLRLNENKLSLASK 860
Qy 1025 NTQTLTADVKONETIEQOKIFSLIOEKNELOQMLSEVIAEKBOQLKTDLKENIEMTINQ 1084
Db 861 ----ACLQDSYDN--LQEIIMKF-----EIDQLSRNLQNPKENETLKSDDLNMEL-LEAE 909
Qy 1085 BELRLGDELKKQOEIVQAEQKHAIKEGELSRCTDLRLAEVEEKLEKESQOLQKQOQLL 1144
Db 910 KE--RNNKLSLOPE--EDKNSK-----EILKVLAEVROEKQKETA 947
Qy 1145 NVQSEMSMOQKINEIENLNKLNKELTLEHMETERLELAQKLNENYEVKSTIKERV 1204
Db 948 KCEQOMAKVQK-----LE-----ESLLATEKV 969
Qy 1205 LKELQKSPETERDHLRGYRIBEAATGLOKTELKIAHLKHEQHTIDELRSVSEKTAQ 1264
Db 970 ISSLEKSRDSKVVADLMNQIQ-----ELRSSVCEKT-E 1003
Qy 1265 IINT--QOLEKSHTKL-----OREIPVHSEOEQLLPVKKVSETOETMNELELLTQEST 1316
Db 1004 TIDTLKQELKDKINCKYNSALVDREESRVLIKQEV-----DILDUKETL-RURILSE-- 1054
Qy 1317 TKDSTTLARIEMERLRNLEKFOESOEIKSITKERNLKTIKEALEVKHDOLKEHIRETL 1376
Db 1055 -----DIERMLCEDLAHATEQLNMLTE-----ASKHSGLLQSAQEEL 1093
Qy 1377 AK-----IQESQKQESQSLNMEKONETTKIVSEMEQFKPSALLRIEIMGLSKRLOE 1432
Db 1094 TKKEALIQELQHK--LNQKKE-----EVEQKKE-----YNFKMRQ 1127
Qy 1433 SHDEMKSVAKEKDDLQRLQEVLOESDQLEKNIKEIVAKHLET--EELKVAHCCLEKOE 1491
Db 1128 LEHVMDSAED-----POSPKTPPHFQHLAKULETQOEIEDGGRASKTSLEH 1175
Qy 1492 TINELRVNLSEKETEISTIQKLEAINDKLQKIOEIEYKEEQQLNIKOISEVQENVNELK 1551
Db 1176 LVTKLNEPREVKNAELRLKQQLREM-ENLRLESQQLLEKNWLL-----QGQDDDK 1226
Qy 1552 QFKEHRKAKDSALQIESKMLE-LYNRLQESQEEIQTIMKEEEMKRVQEAL----- 1602
Db 1227 RQKENSQDQHPDNQQLKNEQESIKERLAKS-KIVEEMLKMKADLEEVQSAALYNKEMECL 1285
Qy 1603 ----QIERDOLKENTKEIVAKMKESQEKYQFLKMTAVNETOEKMC-EIEHLKEQEPETQ 1657
Db 1286 RMTDEVERTQTLES-----KAFQEKELRSKLEEMTEEBERTSQAEMLRKQ----- 1332
Qy 1658 LMLENIETENIRLT--QILHLENLEEMRSVTKERDDLRSEETLKVVERDOLKENLR 1710
Db 1333 --VECLAENGKLVGHQNLHOKIQIVVRLKENVRLAEETEKLRANVFLKEKR 1385

RESULT 5
US-09-723-219-2
; Sequence 2, Application US/09723219
; Patent No. 6391613
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6391613el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/723,219
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human

Qy	3	BEGAVAVCVVRVPLNSREESI-GETAQVYWKTDNNVYIQVDGS-----KSFNFD	50
Db	23	EGDAIKVFVIRIPPAERSGSGADGE-----QNLCLSVLSTSLRLHNSPEPKTTFD	73
Qy	51	RVPHGNETKNVYEIAAPIIDSALQVNGTIFAYGOTASGKTYMVG-----SEDHL	103
Db	74	HVADVDTQSVFAVAKSIVESCWGGNGTIFAYGOTGSGKTYMGPSSDNPFSNLR	133
Qy	104	GVIPRAHDIF-----QKIKFPDRBFLRVSMEIYNETITDLCGTQMKMLPIIREDV	158
Db	134	GVIPRSPXYLFSLIDREKERAGAKSFLCKCSFIEYNEQIYDLL--DSASAGLYLREHI	191
Qy	159	NRNVYVADLTVEVVYTTSEMALKWITKEKSHRYGETQWQSSRSRSHIFRMLIESRKGE	218
Db	192	KGGVVVGVAQEVQVVTSAEAQVLSGWRNRVASTSMNRSSRSRSHAVFTTIESMEK--	249
Qy	219	PSNCEGSKVYSHLNLVDLAGSERAQAQGAAGVRLKEGCNINRSFLILGOVTKKLSGQVG	278
Db	250	SNEIVNRTSLNLVDLAGSERQKDTAEGVRLKEAGNINRSLSCLGVITALVD--VG	306
Qy	279	G-----FNYRDSKULTRILQNSLGNPKTRICITTPVS--PDETLTALQFASTAKYMKNT	332
Db	307	NGKQRHVICYRDSKLTFLRLDSLGNAXTAIIANVHPSRCFGETSLTNFAQAKLIKXK	366
Qy	333	PYNE-----VSTDEALLKRYKEIMDLKK-OLEEVSLETR-----AQAMEKDQLAQL--	380
Db	367	AVNEDUTQGNVSQLOAEVVKRKEQLOELAGSGQTPPEPSFLTRDKKNTYMEYFOEAMLFFK	426
Qy	381	-----BEKDLLQKVONEKIENLTRLMLVTSLSLITLQOELKAKRVRVWCLGKINKMKSNY	436
Db	427	KSQEKKSLTEKV-----TQLEDLTJLKEKFI-----QSNKMIVKPR	463
Qy	437	ADQFNPTNITTKHKLSI-NLIREIDESVCSRSDVFSNTDLTSE-IENWPATKLLNOE	494
Db	464	EDQI-----IRLEKLKHESRGSGFLPEEQDRLISE--LRNEIQTLEQIEHHPVAKYAME	516
Qy	495	NIESELNSRADDYDNLVDYEQLTEKEEMELKIKERKNDLDEFEALERTKKDQEQMLIH	554
Db	517	N-----HSLR-----EENRRLLE-----PVKRAQEM-----	539
Qy	555	EISNLKNLVKRVYVQDLENELSSKVLLERKEDQIKKLOEYIDSQKLENIKMDLSYSL	614
Db	540	---DAQTIAKLEAF-----SEISG---MEKSD-KNQOGE-----	567
Qy	615	ESTEDPKMQKOTLFDFAETVALDQKRESAPLSENLELKEKMK-ELATTYKOMENDILQY	673
Db	568	---SFAQKEPCLPANT-----EKLKAOLLIQIOTELNNSKQEYE	603
Qy	674	SQLEAKKXQVDDLEKELQSAFNEITKLTSLIDGKVPKDLLCNLEBGTITDLQK-ELNK	732
Db	604	EPEKLTKRQOLESELOS-----LOKANINLE	631
Qy	733	--VEENEALREEVILLISELSKSPSEVERIRKETQDKSEELHIITSEKDLFSEVHVKEGR	790
Db	632	NULLEATKACKRQ-----EVSQNLK---IHAETLIKIIITP---TKAYQLHSR	671
Qy	791	VOGLLEIIGTKDDLATTSQNYKSTQDEQFONFKTLHMDPEQYKMWLER---NERNMQEI	847
Db	672	-----PVPKLSPEMGSFGSLYT-----QNSILND-----ILNEPVPPEMNEQAF	712
Qy	848	VNLSKEAQFPDSSLGALKATLSYKTQEOEKTEVEQERLNEMEQLKEQLENRDSPLQTVS	907
Db	713	EATISELRVTQEQMSALQAKI-----DEEHNKLKQOHVDKLEHSTQMOELFS-----	762
Qy	908	REKTLITLQKQTLLEEVKTLITQEKDDLQLOESLQTERDQKSDIHD--TYNNNIDTQSO	965
Db	763	-----SERIDWTQOQFELLSQLNVLEKQLOQET-QTKNDFLKESEVHDLRVLHSDAKE--	813

; ORGANISM: Homo sapiens
TUC-09-014-250-11

Qy	966	LRNALESIKOHOETINTLKSKI SEVS - RNLHMEENTGETKDBFQQQWVGDKKQDLEAK	1024
Db	814	---LSSVKLEYSSFKTQBEKFNKLSERHMHV-----QIQLDNLRLENKLEKLESK	860
Qy	1025	NTQTLTADVXDNBEITIEOQRKIPSLIQEKNELQOMLESVIAEKQOLKTDIKENIEMIENQ	1084
Db	861	---ACLOQSDYN---LQSEIMKP-----EIDQLSRNLQNFKKENETKLSDLNNIMEL-LEAE	909
Qy	1085	BEFLLLGDELKQOQEIIVAOEKNHAIKKEGELSRTCORLAEBVEEKLEKESQOQKOOQLL	1144
Db	910	KE--RNNKLSLQPE---EDKENSCK-----EILKVLAEVRQEKQETA	947
Qy	1145	NVQEWEMGOMKXINEINENKXELKNKELTUEHMETRERLELAQKLNENYEVKSITKERV	1204
Db	948	KCEQMAKQVK-----LE-----ESLLATEKV	969
Qy	1205	LKELOKSFETERDHLRGYIREIBATGLQTKBELKXIAHILKEHOETIDELRRRSVSEKTAQ	1264
Db	970	ISSLEKRDSDKKVADLMNQIQ-----EURLSVCSBKT-E	1003
Qy	1265	IINT--QDLKESHTKL-----QEEIPVLHEQOELLPNVKVSETOBTNMELELLTEQST	1316
Db	1004	TIDTLKQELKDINCKYNSALVDRSESRVLIKQEV---DILDLKETL-RURLISE---	1054
Qy	1317	TKDSTTLARTEMERBLRNEKQFQSEBEISKLTKVERDNLKTIKEALVEKVDQDKHEHRETL	1376
Db	1055	-----DIERMLCEDLAHATEQLNMLTE-----ASKHSGILQSAQEEL	1093
Qy	1377	AK-----IQESQSKOEQSLNMKEDNKTITKIVSMEQFKPKDSALLRIETIEMGLSKRLOF	1432
Db	1094	TKKEALIQELQHK-----LNQKKE-----EVEQKKE-----YNFKMQ	1127
Qy	1433	SHDEMSVAKEKDDLQRLQEVLOSQSDQLENIKEIVAKHLET-EBELKVAHCCLEKEQE	1491
Db	1128	LEHVMSAAED-----PQSPKTPPHQTHLAKLETQEOEIEDGRASKTSLEH	1175
Qy	1492	TINELRVNLSEKETEISTTQKQLEANDKLQNKIQIBIYEKEQOLNIKQISEVQENVNELK	1551
Db	1176	LVTKLADREVEKNAETIRMKQELREM-ENLRLESQOLIEKNMLL-----QGQLDDTK	1226
Qy	1552	OPFHRKAKDSALQSTESKMLE-LTNRLQSOEETIOMIKEEMKRVQOAL-----	1602
Db	1227	RQKNSQDNHPNQOLKNQESIKERLAKS-KIVEMLKMDLVEVQSALYNKEMECL	1285
Qy	1603	-----QIERDQLENTKEIVAKKESQEKSEYQFLKMTAVNETQKMC-EIEHLKEQFETOK	1657
Db	1286	RMTDEVERTQTLES-----KAFQEKQLSKLEEMEYERERTSQEEMLRKQ-----	1332
Qy	1658	NLNEENTETENIRLT--QILHENLEMRSVTKERDDLRSLVREETLKVERDQLEKNLR	1710
Db	1333	--VECLAENGKLVGHQNLHOKIQYVVRUKENVRILAEETEKLRANVFLPEKKR	1385

Query Match 7.9%; Score 1052.5; DB 4; Length 3878;
 Best Local Similarity 19.5%; Pred. No. 2.2e-41;
 Matches 669; Conservative 603; Mismatches 1069; Indels 1089; Gaps 144;

Qy 26 TAQVYKTDNNVIVQDGSFNFDRVPHGN-----ETTKNVEEIAPIIDSAIQGYNGT 81
 Db 307 TMOISFLOEKIVYMEQDK-----KVENSKBEIQEKETIEELNTKIIE----- 353

Qy 82 IFAYGQTASGTYTMMG-----SEHGLGVIPIRAIHDFQIKIKFFDPRELLAVSYMEIYN- 136
 Db 354 -----EKKTELKDKLTTADKLLGLBLOEQIVQKNOEIKNMK-----LELTNS 395

Qy 137 -----ETITDILLCGTKKPLIREDVNRNVVADLTBEVVYTSBALKWITKGEKS 188
 Db 396 KOKERQSSEETIKQMGVTEELQ-----KRNHDKDSQFETDIVQMEQETQKLEQL-RAELD 450

Qy 189 RHYGE--TKMNQRSRSHTIFRMILESREKGEPSNC-----EGSVKVSHNLVLDL 236
 Db 451 EMYGOQIVQKQELIRQHMAQWEEKMRHKGEMENALRSYNSITVNEQIKLMVAINEL 510

Qy 237 -----AGSERAAQTGAAGVRLKEGCNINRSILFILGQVKKLSLDG-----QVGGFINVRD 285
 Db 511 NIKLODTSQKEKLKBEELGLILEKCALORQ-----LEDVLEELSFREQIARQRTIAEQE 567

Qy 286 SKLJTRILQNSLGGNPKTRIICTITP-----VSFDETLTALQFASATKYMKNTPYVNEVST 340
 Db 568 SKL-----NEAHKSLSTVEDLKAEIVSASERKELE-----LKHEAEVTNYKI 610

Qy 341 DBALKRYRKEIMDLKKQLEVSLE---TRAQMEKQDQAOLLEKSKLLQKVONKEIENL 397
 Db 611 KLEMLEKKNVLDRAESQEAELERLTQLLFSHEEELSKLEDLEIHRINIEK----- 666

Qy 398 TRMLVTSSSLTLOELKAKKRVVTCWLGKINKMNSVADQFNPTNITTKTK-LSI 455
 Db 667 -----LKDNLGIHYKQOIID---GLQEMSQKIETMQFE-KDNLIITQONQILLEI 711

Qy 456 NLLREIDES-VCSESDFVSNTLDTL-SETEMNPATKILNQENIESELNSLRADYNLVDL 513
 Db 712 SKLQLOQLVNSKSEEMTLOINELQKEIE-----ILAQE-EKEKGLTEQEVQEL- 760

Qy 514 YEOLTEKEEMBELKKEKNDLOB-----KQOEMQLIHEISNLKLV---KHREYVNDLE 574
 Db 761 --QLATLELEKQWKEKE-NDLOEKAQFAEANSILKDEKKTILEDMLKHTPVSQERLIF 817

Qy 537 FEALERTK-----KQOEMQLIHEISNLKLV---KHREYVNDLE 574
 Db 818 LDSIKSKSDSVWEKEIEILIBENEDLKQOQIOLNEEIEKQRTTFSAEKNFEVNYQELQ 877

Qy 575 NELSSKVELLREKEDQIKKLOEYIDSKLENIMKMDLSIESIEDPK-QMKQTLFD- 629
 Db 878 BEYACLLKVKDLEDSD-KNKQELLEYKSKLALNEEL--HLQINPTTVKMKSSVFDEKDT 934

Qy 630 --AETVAL-----DAKRESAFURSENLEKKEKELATYKQWENDIQLYQSOLEA 678
 Db 935 FVAETLEMGVEVKEKDTTELMEKLEVTREKLELSQRLSDLSQELQKQKHGEISFLNEBVS 994

Qy 679 KK-KMQVDLE-KELQSAFNE-----ITKLTSLDGKVPKOLLNLELGGKITDLQ 726
 Db 995 LKQKEQVLSRCELEIINHNAENAVQSCDTQVSSLLDGV--TWTSGAGSVSKYN 1051

Qy 727 KELNKE-----VEE-----NEALREEVILLSSELKSL-----PSEVERLRKEI 763
 Db 1052 KSGEESKIMVEDKVSFENNTVGBESQEQILDLPLPSVTKESSLRATQPSNDKLQK-- 1109

Qy 764 QDKSEELHIITSKKDLKLFSEVVIKESRVOGLLEIEIKTKDLDLATQSNYKSTQEQFQNF 823
 Db 1110 -----ELNVLSKQNDL-----RLQMEAORI-----CULSVYST----- 1138

Qy 824 TLHMDPEOKYKWLLEENRMNQEIVNLSKEAQFDDSLGALKTELSTYKTOELQEKTRVQ 883
 Db 1139 --HVDQVREYM-----ENEK-----DKALCSLKEELIFAQ---BEKIKELQ 1174

Qy 884 ERLNEMQLKEQLNDRDSPLOTVEREKT-----LITEKLOOTLEB-----VKILTQE 930
 Db 1175 -KIHOLE-----LQTMKTOETGDEKGLHLHIGLKQKAVSECECYFLOTLCV 1221

Qy 931 KDDLKQLQESLQIERDQKSDIHDTVMNIDTQBLRNALBSLKHQOFTINTILSKISEE 990
 Db 1222 LGYEYTPALKCEVNAED-KENSGDYISENED--PELODYRYEVQDFQNMHTLLNKVTEE 1278

Qy 991 VSRNLHM-----FENTGETKDFQKQVGVGIDKKQDLKAKNTQILT 1030
 Db 1279 YNKLLVLQTRLSKIWGQOTDGMKLEFGEENLPKEETEFLSIHSQMTNLEIDVNHKSLS 1338

Qy 1031 A-DVKNWEIIEOQKPIESLI---QEKNELQOMLES-----NOBEL----- 1061
 Db 1339 SLODEKTKLEQVQOELESLSLQOQKLETEQNYAEIHLCKLQRLQAVSESTVPPSLPV 1398

Qy 1062 --VIAEKEQLKT---DLKENIEMTIE-----NOBEL----- 1087
 Db 1399 DSVVITESDAORTMYPGSCVKKNIDGTTFEFGFVKEETNIVKLEKQYQOLBEEVAK 1458

Qy 1088 -----RLLGDE---LKKQOEIVAOEKNAHKKGBELSRICDRLA--- 1123
 Db 1459 VIVSMSIAPAOQTELSRISGGKENTASSKQAHAVQOQBHYF---NEMKLSQODIGFQTF 1515

Qy 1124 -EVEEKLKESQOQ---EKOQQLNVQEESEMSEMKKINETE 1161
 Db 1516 ETVDYKFBKBPPLSKSELGEHGKEILLNSDPHDIPESKDCVLTISEEMFKDKTFFVRQ 1575

Qy 1162 NLKNEKL-----NKLTEHMETERLELAQKLNENYEVKSITKERVKLKELQSKFETE 1215
 Db 1576 SIHDEISVSSDASQMLNHEOLE--DMRQELVRQYQEQHQATQSSIDNE--NLVSE 1630

Qy 1216 RDHLGYRTREIATGLQYKELKIAHIHLKHEQETIDELRRSV-----SEKTAQIINTQ 1269
 Db 1631 RE-RVLLBELEA-----LKQLSLAGREKLCCELRNSSTQNGENQGEVEBOT 1678

Qy 1270 DLEKSHTKLEHIEP-VLHEEQE-----LLPNVKKVSETQETMNE--LELLTE 1313
 Db 1679 FKEKELDRKPEDVPPEILSNERYALQKANNRLKILLEVVKTTAAVEETIGHVGLDRL 1738

Qy 1314 QSTTKDSTTL-----ARI-----EMERLRIN-----EKF 1337
 Db 1739 SSKSQSSASLIWRSEASVSKVCHHEHTRVTDSPSYSGSDMPRNDINMWSKVTEGT 1798

Qy 1338 QSQSEIKS-----LTKERDNLK-TKEALEVKHDOLKEHIRETILAKTQESQSKO--- 1386
 Db 1799 ELSQRLVRSFGAGTIDPENEEIMLNLISRLQAQAVEKLELLEISETSSQLEHAKVTOTELM 1858

Qy 1387 EQSLANKEKDNETTIVSEM-----EQLK 1410
 Db 1859 RESFRQKQATSELKQCEELRHLHEESPARQLAVELSKAGVIDGYADEKTLPERQIQ 1918

Qy 1411 PKDSALLRIEIMLGLSKLOESHDMSKVAKEKDDLOBLQELVLOSES-----DOLKENIK 1466
 Db 1919 EKTDIIDRLQELLCASNLQLEAEFQOQIQEERELLSRQKEAMKAGPVEQOLLQETE 1978

Qy 1467 EIVAKHLET-----EELVAKHCLCQEBETINELRVNLSEKETEISTIOKOLE 1515
 Db 1979 KLMKEKLVQCAKAEVRDLDQKQKALEIDVEEQSVSRFIELE---QEKTELMDLQOQ 2035

Qy 1516 AINDKLQNK-----IOEIVEKEEQLMI---KOISEVQENVNELKQF 1553
 Db 2036 ALEKQLEKMRKFTDEQAIDREHERDVFQEIQKLEQQLKVVPRFQFISEHQ--TREVEQL 2093

Qy 1554 KEHRKADSALOSIESKMLELTLNRLQESOEI-----QIMKEKEEMKRVQE-- 1600
 Db 2094 ANHLKEKTDKCSLLSKQLOQORDIQERNEEIEKLEFRVRELEQALVSDATFKQVDRK 2153

Qy 1601 -----ALOJTERDQKENTYKEIVAKMESQEKYQFLKMTAVNETQEKWCE 1645
 Db 2154 HFGAVEAKPELSLEVLQQAERDAIDRKEKIT---NLEQLQEPRE-----ELEKNEE 2204

Qy 1646 IEHLKEQOFQK-----INLENITENIRITQILHENLEEMRSVTKERDRLRSVEETL--- 1698

[illegible]

Db	3234	LEQKQQLLNESQKIESQRMLYDAQLSEE-----QGRNLEL-----QVLLSESE	3276
Qy	2511	HIRLEKEISIKLQKQNE-----OLIKOKNELLNNQHLSNE	2545
Db	3277	KVTRREMSSTLDRERELHAQLQSSDGTGQSRPPLPSDDLKELQKQLEEKHSRIVELLNE	3336
Qy	2546	VKTWKERTLK-----REAHKQVTCENGSPKSPKVTGTGATSKKQKITPSSQCKERNLQDPV	2597
Db	3337	TEKYKLSLQTRQOMKXDRQVHRKTL-----QTEQEANTEGQKK-MHELQSKVEDLQRL	3390
Qy	2598	PKESPCKCFDPSRSKSLPSPHPVRYFNNSLGLCPQVQNAQESVD-----SQ	2645
Db	3391	BEKQYQVYKLDLEGORLQ-----GIMQEFQKQELEREKREKRSRRILYQNLNE	3437
Qy	2646	PGPWHASSGK	2655
Db	3438	PTTWLSLTSR	3447
RESULT	7		
US	08-353-700-1		
;	Sequence 1, Application US/08353700		
;	Patent No. 5599919		
;	GENERAL INFORMATION:		
;	APPLICANT: YEN, TIMOTHY J.		
;	APPLICANT: RATTNER, JEROME B.		
;	TITLE OF INVENTION: NUCLEIC ACID ENCODING A		
;	TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,		
;	TITLE OF INVENTION: AND METHODS OF USE		
;	NUMBER OF SEQUENCES: 4		
;	CORRESPONDENCE ADDRESS:		
;	ADDRESSEE: DANN, DOREMAN, HERRELL AND SKILLMAN		
;	STREET: 1601 MARKET STREET, SUITE 720		
;	CITY: PHILADELPHIA		
;	STATE: PA		
;	COUNTRY: USA		
;	ZIP: 19103-2307		
;	COMPUTER READABLE FORM:		
;	MEDIUM TYPE: Floppy disk		
;	COMPUTER: IBM PC compatible		
;	OPERATING SYSTEM: PC-DOS/MS-DOS		
;	SOFTWARE: PatentIn Release #1.0, Version #1.25		
;	CURRENT APPLICATION DATA:		
;	APPLICATION NUMBER: US/08/353,700		
;	FILING DATE: 09-DEC-1994		
;	CLASSIFICATION: 435		
;	ATTORNEY/AGENT INFORMATION:		
;	NAME: REED, JANET E.		
;	REGISTRATION NUMBER: 36,252		
;	TELECOMMUNICATION INFORMATION:		
;	TELEPHONE: (215) 563-4100		
;	TELEFAX: (215) 563-4044		
;	INFORMATION FOR SEQ ID NO: 1:		
;	SEQUENCE CHARACTERISTICS:		
;	LENGTH: 3248 amino acids		
;	TYPE: amino acid		
;	STRANDEDNESS: single		
;	TOPOLOGY: linear		
;	MOLECULE TYPE: protein		
;	HYPOTHETICAL: NO		
;	ANTI-SENSE: NO		
;	ORIGINAL SOURCE:		
;	ORGANISM: HUMAN		
US	08-353-700-1		

```

Query Match      7.8%; Score 1040.5; DB 1; Length 3248;
Best Local Similarity 19.6%; Pred. No. 6.7e-41;
Matches 635; Conservative 583; Mismatches 1033; Indels 985; Gaps 132;

Qy 154 IREDVNRNVYADLT EEVYVTSEMALKWTTTKEGRHYGETK----MNRGRSRSHITFRM 209
||| : : : : :
| : : : : :
| : : : : :
| : : : : :

Db 237 IRRDFSAFYFSGELE---VTPSRSTIQIGKRDANSFFGNSSPHLLDQKQNOELRNK 293
||| : : : : :
| : : : : :
| : : : : :
| : : : : :

```


QY 210 I--LESREKGPNSCEGVS-KVSHNL-----VDLAGSER-----AAOTCAA 248
 Db 294 INELERLQHEKEMKGVNFQBLQLEKAKVELIEKERVANKRCDLVRTTAQYDQA 353
 QY 249 GVRUKBGCNINRSFILGOVTKLSDGVGVFINVRSKLTRIILONSIGMGPKTRIICTI 308
 Db 354 STKYT-----ALEQKLKLTEDLSCORQNAESARCS--LEQKIKEXEK-----394
 QY 309 TPVSFDETLTALQ-----PASTAKYMKNTPVVNEVSUDE--ALLKRYRK 350
 Db 395 ---BFOBELSQOORSFQTLDOBCIQMKARLTQELQOAKMNVLQAEIDLKLTYSVKQOLEN 451
 QY 351 EIMDLKKOLEVSLTRAQAMEKQDLAQLLE-----EKDLQKVQNEK-----IEN 396
 Db 452 NLEEFKQKLCRAEQAFQASQIKENELRSMEEMKKNLLKSHSEQAKREVCHLEAELKN 511
 QY 397 LTRMLVTSSITLQOELKAKRRVTVCLG---KINKMKNYADOFNIPNITTKTHKL 453
 Db 512 IKQCL--NOSQNFABEMKAKNTSOETMLRDLQEKINOQOENSLTLEKLAIVADLEKQDRC 569
 QY 454 SINLLREIDSVCSGESDVFSNT---LDTLSEIE-----WNPAT-KL 490
 Db 570 SQDLKKGHEHIEQNDKLSKTEKESKALLSALELKKEVEELKEEKTLPSCWKSENEKL 629
 QY 491 LNQ-----ENTIESELNSURA-----DYDNLVLDYQRLTEKEEMELKLEKNLDL 535
 Db 630 LTOMSEKENLQSKINHLETCLTKTOQIKSHEYNERV---RTLEMDRENLSVEIRNLHVL 686
 QY 536 EPEALERKTKK--DOEMQLIHHSNMLKVLKREYVNOQDLENELSSKVELLREKEDQIK 593
 Db 687 DKSVEVETOKLAYMELQAKAEFSQK-----HKEIENMWLKTSLQTLQGVEDLEHK 738
 QY 594 LQ-----EYIDSKLBNRMWLSYSESTEDPKOMKOTLFDABTV-----ALDAK- 638
 Db 739 LQLLSNEIMDK--DRCQDLHAESYERLCLKSK-----DASLVNEDHORSLLAFDQOP 791
 QY 639 -----RESAFRSENLEKEMKELATTYKOMENDIQYOSQLEAKKQMVULEK 688
 Db 792 AMHGSFANIGQSGMPSERSECRLEADQSPKNSAILQNRVDSLEFSLESQKMSDLQK 851
 QY 689 ELQSAFNEITKLTSLIDGVKPDLL-----CNLEEGKITDLOKEL-----NKEVEEN 736
 Db 852 QCE-----ELVQ-----IKGEIENLMKABQMHSFVAETSQRISKQEDTSAHQNVVAETL 903
 QY 737 EAL-----BREVILLSELKSPLEVERLRKE---IQCKSELHIITS-----EKDLKFSEV 785
 Db 904 SALENKEKELQILNDKVEQAEIOELKSNHLLSDLSKELQELSETLSLEKKEMSSIIS 963
 QY 786 HKESRVQGLLEIGTKD---DLATTQSNYKSTDOEQFQNFKTLHMDFOQKYQWVLENER 842
 Db 964 LMKREIEELTQENGTLKEINASINQEKMNLIQKSESFANY-----IDEREKSISELSQ 1017
 QY 843 MNOEIVNLSKEAQKFDSSIGALKTELSTYKTOELOKTRV-----QERLNEBQ 891
 Db 1018 YKOEKLILQRCBE---TGNAYEDLSQYKAAQEKNSKLECLNNECTSLCENRKNLEQ 1073
 QY 892 LKEQLENRDSPIQOT-----VEREKTLI--TEKLQOTLEEVKTLTQ-----EKDDLKQLOE 939
 Db 1074 LKEAFAKEHOFELTKLAFABERNQNLMLLETVQALRSEMTDNQNSSEAGGLKQETM 1133
 QY 940 SIQIERDQKSDIHDTPVNNITDQELRNALSELKQHQETI--NTLKSQISE-----989
 Db 1134 TLKEEQNKQKEVNDLLQNEQELMKVMKTKHECONLESEPIRNSVKRESERSONCQNFQ 1193
 QY 990 -----EVS-----RNLMHEENTGETKDEFOQKVVGDIKQDLEAKNT 1026
 Db 1194 MLEVKEISLSDSVNALQVQLEAMLRNKLKQISEKE--KECLOHELOTI--RGDLETSNL 1250
 QY 1027 QILTAD-----VKDNETIEQOKRIF-----SLIQEKNELQQMLSEV 1062
 Db 1251 QDWQSOEISGLKDCIDEABEKYISGPHELSTSQNDNAHLQCSLQTTMKNLEKEICEIL 1310
 QY 1063 IAEKEQLKTLKEN-----TBMVTIENQBELLLGDELKKQOEIVAQEKHAIKKE-----1112

Db 1311 QAEKYELTELNDRSSECITATRKWABEVGKLLNEVK-----ILNDSSGLLHGLVEDIP 1365
 QY 1113 -GELSRCTDR-----LAEVEEK-----LKEKSQQLQ--EKQOQLLNQVEE-----1149
 Db 1366 GGEFEGEPNEQHPVSLAPLDESNSYEHLLTSDKEVQMFHAELOEKFLSLQSEHKILHQH 1425
 QY 1150 -----MSEMQKLTNEIE-----NLKN-----ELKNKELTLEH-----1176
 Db 1426 QMSSKMSSELQTYVDSLKAENLVLSLNRNFGDLVKEMQLGLEBGLVPSLSSSCVDPSS 1485
 QY 1177 -----AQL-----METERLEL 1184
 Db 1486 SLSSLGSSSFYRALLEOTGDMSSLLNLEGAVSANQCSVDVFCSSLYQYVDSLKAENLV 1545
 QY 1185 AQKL-----1188
 Db 1546 STNLNFGDLVKEMQLGLEBGLVPSLSSSCVDPSSSLSSLGDSFYRALLEOTGDMSSLL 1605
 QY 1189 -----NENYEEVKSITKERVKVLKELQKSFETERDHLRGYIRE 1225
 Db 1606 SNLEGVVSANQCSVDVFCSSLOEENLTRKETPSAPAKGVBELESCEVYQSLEKLEBK 1665
 QY 1226 IEATGLQTKBELKTAHILKHOETIDELRR-----SVSE 1260
 Db 1666 MESQIMKNKEIOELEQLLSERQELCLRKQYLSENEQWQOKLTSVTLMEKSLAAEKK 1725
 QY 1261 KTAQI-----INTQ-----DLEKSHTKLOEIPVUHE 1287
 Db 1726 QTEQLSLEVARLQGLDLSRSSLLGIDTDAIQGRNESCDSIKESHTSETTERTPKHD 1785
 QY 1288 -----EQELLPNVKVKSET-----1301
 Db 1786 VHQICDKDAQODLNDIDIEKITETGAVKPTGBCSGEQSPDTNYPPGEDTKQGSSECSIEL 1845
 QY 1302 -----QETMNELELLTQOSTTKDSTTL-----A 1324
 Db 1846 SFGSNALVPMDFLGNQEDIHNLQVRVETSNENRLLHVIEDRDRKVESLLENMKEKLD 1905
 QY 1325 RIEMERLRLNKFO---ESOEIISLTKBERDNLKTIKEALEVKHQL-----1368
 Db 1906 KLHLQEVQWTKIEACIELEKIVGELKKENSLSLEKLEVFSCDHQELLQVRVETSEGLSD 1965
 QY 1369 -----KEHIRTAKIQUESQKQSLNMKEKNETTKIVSEMEOFKPKDSALLRI 1419
 Db 1966 LEMHADKSSREDIGNVAKVNDSW--KERFLDV---ENELSRIRSEKASIE--HEALYLEA 2019
 QY 1420 BTEMGLSKRLQESHEDEMK-----VAKEKDDQLQRLQVLOSES---DQLKEN 1464
 Db 2020 DLEVVTQTEKLCLEKNENKQVIVCLBELSVVTSERNQLRGELDTMSKTTALDQSEK 2079
 QY 1465 IKEIVAKHLETEEE---LKVAHCLKQEOETINELRVLNSELKETEISTIOKLEAIND 1519
 Db 2080 MKE-KTQELSHQSECLHCIOVAEAVKTEKTELLOTLSSDVSELLAKDKTHLOKQLEK 2138
 QY 1520 -----KLONKIOEYKEEQLNIQISEQVQNVNVELQFKEHRAKOS--ALQSLIE 1568
 Db 2139 DSQALSLLTKCELENOIAQL--NKEKELLVKSESLOARSE---SDYEKNVSKALEAL 2193
 QY 1569 SKWLELTLNRLOESQEIQIMIK-----EKEEMKRVQOEALQI-----ERDOLKENTPK 1614
 Db 2194 VKGGEFALRLSTQEVHQLRGIEKLRVRIEADKKQLHIAEKLERENDSLDKKVE 2253
 QY 1615 EIVAKMKSQEKYOFKMTAVNTOEKRCRIEHLKQEPETO---KMLNLETETEN 1667
 Db 2254 NLERELQSEENQ-----ELVILDAENSKADEVETLTKTOEEMARSILKIFELDLVTIRSEK 2308
 QY 1668 IRLTQILHEN-----LEEMSVTKERD--DLRSVEET---LKVERDOLKENLRET 1712
 Db 2309 ENLTQIQIQEQOGLSDELKLLSSFKSLEEKQAEIQIKESKTAVENTLQNLKE--LNEA 2367
 QY 1713 ITRDLEKQBELKIVMHML-----KEHQ--ETIDKLRG--IVSEKTNISNNQ--KOLEHSN 1762


```

Db 2368 VAALCGDQOEIMKATEQSLDPPIEEHQLRNSIEKLRARLEADEKQKQCVLQQLKSEBHA 2427
QY 1763 DALKAQDLKIQEELRIAHMLKQOEITDKLGVSEKTKLNMQKDLNSNAKLOEKI 1822
Db 2428 DULKGRVENLERLEIA-----RTNQ-----HALEAENSKEVEITLAKIEGMT 2473
QY 1823 QELKANEHQLITLKD-----VNETQK-----VSEMEQLKKQIKDOQL 1861
Db 2474 QSLRGLELDVVTIRSEKENLTNLOEQERISELEIINSFENILOEKEQKQVOMKESS 2533
QY 1862 TSLKL---ETENLN-LAQELHENLEEMKSVKMDNLRVRVEITLKLRLDQLKESLQTKA 1917
Db 2534 TAMEMLQTLKELNERRVAALNDQEA---KEQNLSSQVECELEKAKAQLLOGLDEAKN 2590
QY 1918 RDLIEIQELKATARMLESKEHETVDKLEKISSEKTIQISDKDLDKSD-----1966
Db 2591 NYIVLOSSVGLLOFVEDGKQLEKDELSRLKQIQDOEQVLVSKLSQVGEHQLWKEQ 2650
QY 1967 -----ELQKIQELQKQELQLLRVKEDVNMHKKI-NEMEQLKKQPEPNYLCKEM 2016
Db 2651 NLELRNLTVLEQKIQVLOSKNASLODTLEVLOSSYKNLENELELT-----M 2698
QY 2017 DNFQITKLHLESEETRIVAKERDELRRIRKESLMDRDQFIATLRMIARDRONHVKPE 2076
Db 2699 DKMSFVEKVN-----KMTAKETELQEMHEMA-----QKTABLE-----ELSGE 2738
QY 2077 KRLLSDGQOHLMESLRKESKRIKELLKRYSEMDDHVECLNRLSLDLKEIELEFHRIMKGLK 2136
Db 2739 KNELAGELQILLLEIKSSKQDLKELTLENSELKKSJDCMHK--DOVEKE-----2785
QY 2137 YVLSVYTKIEBQHECINKFEMDFIDEVEKQELIKIHLQODCDVPS-RELRLDLKLNQ 2195
Db 2786 -----GKVREE-----IABVQLR-LHEAEKQHALLDNTNKQVEIQTVE-----KLIS 2830
QY 2196 NMDLHTEELKQSESEFFSIKTEFOVLNKRKEMTOFLEWLNTRFDIEKLKNGIKEN 2255
Db 2831 K-----ECLSS-QKLEIDLKSSKSELNLSKATQIILEELKTKTMDNLKYVNLKKN 2884
QY 2256 DRICQVNNFFNNRIIAIMNESTEFERSATISKEMFQDLKSLKEKNEKLPKNVQTLKTSL 2315
Db 2885 ERA-----QGMKLLIKSKQLEKEELKQELKLSQ-LQAQEK-----2921
QY 2316 ASGAQVNPPTQDNKNPHVTSRATQLTKERILENSLHEAKESA---MHKESKIKMQKE 2372
Db 2922 ----QKTGTWMDTK-----VDELTE-IXELKETLEKTEADRYLKYCSLLISHEK 2969
QY 2373 LEVTNDIIAKQAVHESKCLIEKTEITQ--VLQKVAGLAKPYKEETED-LMKLVKI 2429
Db 2970 LEK-----AKEMLETOVHLCSQSQSDRGSPFLGFPVPGSPIPSVTEKRLSSQNK 3024
QY 2430 DLEKMKNAKEFEKEISATKATVE-YQKEVIRLLRENRLRSQQAQDTSVSEHTDPOPSNK 2488
Db 3025 SKQRQSSGIWNGRGPTATPESPSKSKKAVMSGI---HPAEDT-----EGTEFEPEGL 3077
QY 2489 PUTCGGSG-----IVONTKALILKSHIRLEK-EISKLKQONPOLIKQKNELIS 2537
Db 3078 PEVVKKGFPADIPGKTSPYILRTTWMATRTSPRLAAQKALSPLSGKENLAESSKPTAG 3137
QY 2538 NNQHLNNEVKTWKERTIKREAHQVTCENSPSPKVTGTASKKKQITPQCKERNLQDPV 2597
Db 3138 GSR---SQKVKV-----AQSPVDSGTILREP---TTKSPVFNLPERS 3175
QY 2598 PKESPCKSCFFDSKSLSPHPVRVFDNSSLGLCPVQVAGAESVDSQGPWHASS 2653
Db 3176 PTDSREGRLVRGRVLPAK-----LDWSQLAV-RTVRSSEALCVSD---PWEVQS 3223

```

RESULT 8

PCT-US95-16216-1

; Sequence 1, Application PC/TUS9516216

; GENERAL INFORMATION:

; APPLICANT: Yen, Timothy J.

; APPLICANT: Rattner, Jerome B.

```

; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100

```

```

; TELEFAX: (215) 563-4044

```

```

; INFORMATION FOR SEQ ID NO: 1:

```

```

; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 3248 amino acids

```

```

; TYPE: amino acid

```

```

; STRANDEDNESS: not relevant

```

```

; TOPOLOGY: not relevant

```

```

; MOLECULE TYPE: protein

```

```

; HYPOTHETICAL: NO

```

```

; ANTI-SENSE: NO

```

```

; PCT-US95-16216-1

```

Query Match

Best Local Similarity 7.8%; Score 1040.5; DB 5; Length 3248;

Matches 635; Conservative 583; Mismatches 1033; Indels 985; Gaps 132;

```

QY 154 IREDVNRVYVADLVEEVYVTSEMALKWITKESKSHYGETK-----MNQSRSRSHYTFM 209
Db 237 IRRDFSASYFGELE---VTPSRSTLIQIKGRDANSFFGNSSPHLLDQKQNOELRNK 293
QY 210 I--LESREKGEPSNCEGSV-KVSHLNL-----VDLAGSER-----AAQTGA 248
Db 294 INELRLQGHKEKMGQVNFQELQLEKAKVLEIEKVLNKRDELVRTTAQDOA 353
QY 249 GYRLKEGNNINSLFILGOVVKLSDGVGGFINYRDSKLTILQNSLGNPKTRICTI 308
Db 354 STKYT-----ALEQKAKLTEDLSQONAESARCS--LEQKIKEK-----394
QY 309 TPVSFDEITALQ-----FASTAKMNTPTVNEVSTDE--ALLKRYRK 350
Db 395 ---EFOEELSQRQSRFOTLDQECIQMKARLTQELQAKNMHNVLAELDILTSVKQLEN 451
QY 351 EIMDLKKOLEVSLSTRAQAMEKQALOLE---EKDLQKVQNEK-----IEN 396
Db 452 NLEEFQKLCRAEQAFQASQIKENELRMEEMKKNLLKSHSQAREVCHLEAEKKN 511
QY 397 LTRMLVTSSTLTQOELKAKRRVTCWLG---KINKMNSNYADQFNIPNTTITTKHKL 453
Db 512 IKQCL--NQSQNFABEMKAKNTSQETMLRDLQEKINQQENSILTLEKLAVADLEKQDC 569
QY 454 SINLLREIDESVCSSEDFVSNV---LDTLSEIE-----WNPAT-KL 490
Db 570 SQDLLKQREHIEQLNDKLSKTEKSKALLSALEKKEEYELKEEKLTFSCWKSENEKL 629
QY 491 LNQ-----ENTESELSNLR-----DYDNLVLDVEQLRTEKEEMELKLEKNDLD 535

```

Db 630 LTQMESEKENLQSKINHELTCLKTQOIKSHYNERV---RTLEMDRENLSVEIRNLHVL 686
 Qy 536 EFEEALERTTK--DOEMQLIHEISNLKVLKREVVYNQDLENEISSKVELLREXEDQIKK 593
 Db 687 DSKSVETOKLATNEIQKAEFSDQ-----HOKIEENMCLKTSQLTQGVEDLEHK 738
 Qy 594 LQ-----EYIDSKLENKMDLSYLESIEDPKOMKOTLFDATV-----ALDAK- 638
 Db 739 LOLLNSNEMDK--DRCYQDHLAEVESLRDLKSK---DASLVNEDHORSLLAFDQOP 791
 Qy 639 -----RESAFLSENLELKEKMKELATTYKQEMENDQLQVSOLEAKKMQVDLEK 688
 Db 792 AMHHSFANIIGEOGSMPSERSECRLEADQSPKSAILLQNRVDSLEFSLESOKMNSDLQK 851
 Qy 689 ELQSAFNEITKLSIDGKVPKDLL-----CNLELEGKITDLOKEL-----NKEVEEN 736
 Db 852 QCE-----ELVQ-----IKGEISENLMKABQMHSQVAFETSQISKLOEDTSAHQNVVATL 903
 Qy 737 EAL-----REEVILLSKSLPSEVERLKE---IQDKSEELHIITS-----BKDLKFSEV 785
 Db 904 SALENEKEKELQLLNDKVETEQAIEQELAKSNHLLSDLSKELQLLSETLSLEKEMSSIIS 963
 Qy 786 HKRSRVOGLLEIBGKTQD--DLATTQSNYKSTDOEFONFKTLMDFKQKVMVLEENER 842
 Db 964 LNKREIEELTQENGTLKEINASLNOEKVNLIQKSESFANY-----IDEREKGISELSDQ 1017
 Qy 843 MNORIVNLSKEAQKDFDSIGALKTEL SVKTOELQKTRV-----QERLNEMEQ 891
 Db 1018 YKOEKLIILQRCSE---TGNAYEDLSQYKAAQEKNSKLECLNNECTSLCENRKNLEEQ 1073
 Qy 892 LKEQLENRDSPIQT-----VEREKTLI---TEKLOQTLFEVKTLTQ-----EKDDLKQLOE 939
 Db 1074 LKEAPAKEHQEFTKLAFAEERNQNLMELETVQALRSEMTDNQNSKSEAGLQKQIM 1133
 Qy 940 SLQIERTQKSDIHTVANNIITQOLRNALLESKHOETI--NTLKSILSE-----989
 Db 1134 TLKBEQNKQKVEVNDLQENQELMKVMTKQHCQNLSEPIHNSVKRESERNQCNFKPQ 1193
 Qy 990 -----EVS-----RNLMHEENTGETKDFQKQKMGVGDKKQDLKAKNT 1026
 Db 1194 MDLEKVESLDSYNALQVLEAMLRNKNELKLOESEKE--KECLOHELOIT--RGDLTSLNL 1250
 Qy 1027 QTLTAD-----VKONBIIEQQRKIP-----SLIOEKNELOQMLESV 1062
 Db 1251 QDMQSOEISGLKDCIDAEKVIYSGHLSSTQNDNAHLQCSLOTMTMKNLEKICEIL 1310
 Qy 1063 IAEKQKLTDLKEN-----IEMTIENQOEELRILGDELKQOEIVAQEKHAIKKE-----1112
 Db 1311 QAEKVELVTELNDRSSECTATATKMABEVKGILLNEVK-----ILNDDSGLLHGEIVEDIP 1365
 Qy 1113 -GELSRTCDR-----LAEBEEK-----LKEKSQOLO--EKQOQLLNVQEE-----1149
 Db 1366 GGEFGEQPNQHPVSLAPIDESNYSYHETLSDEKQVOMHFAELOEKFLSLQSEHKILHDQH 1425
 Qy 1150 -----MSBMQKINBIE-----NLKN-----ELKNKELTLEH-----1176
 Db 1426 CQMSSKMSSELOTVVDLSKABNLVLSNLRNFQDVLKEMQGLGEBGLVPSLSSSCVPDSS 1485
 Qy 1177 -----METERLEL 1184
 Db 1486 SLSSIGDSSFYRALLEQTGMSLLSNLEGAVSANQCSVDEVFCSSLOTTVVDLSKABNLVL 1545
 Qy 1185 AQKL-----1188
 Db 1546 STNLRNFQDVLKEMQGLGEBGLVPSLSSSCVPDSSSLSSGDSFYRALLEQTGMSLL 1605
 Qy 1189 -----NENYEVKSTITKERVKVLKELQKSFETERDHLRYIRE 1225
 Db 1606 SNLEGVVVSANQCSVDEVFCSSLOEENLTRKETPSAPAKGVEELESCEVYRQSLKLEEK 1665
 Qy 1226 IEATGLOTKEELIAHLKHOETIDELR-----SVSE 1260
 Db 1666 MESQIGMKNKBIQELQELSSERQELDLKQKYLSENEQWQKLTSTVLEMESKLAEEK 1725

Qy 1261 KTAQI-----INTO-----DLEKSHTKLOEIPVLE 1287
 Db 1726 QTEQLSLEVARLOQLDLSRSLGIDTDAIQRNESCDSIKESHTSETTERTPKHD 1785
 Qy 1288 -----BQELLPNVKVKYSET-----1301
 Db 1786 VHQICDKDAQODLNDIEKTIETGAVKPTGECSEQSPDNYEPFGEDKTQGSSECISEL 1845
 Qy 1302 -----QETMNELELLTEQSTTKDSTTL-----A 1324
 Db 1846 SFGSPNALVPMDFLGNQEDIHNLQLRVKETSNENLRLLHVIEDRDRKVESLINEMKELDS 1905
 Qy 1325 RIEMERLRLNKFO---ESQEEIKSLTKERNLTKIKEALEVKHQD-----1368
 Db 1906 KLHLQVQLATKIEACIELEKIVGELKKNESDLSEKLEIFYFCDHQELLQORVETSEGLNSD 1965
 Qy 1369 -----KEHIRETLAKIOESQSKOBOSLNMMKEDNETTKIVSEMQFQPKDSALLRI 1419
 Db 1966 LEMHADKSSREDIGDNVAKVNDWM--KERFLDV---ENELSRIRSEKASIE--HEALYLEA 2019
 Qy 1420 EIMLGLSKRLQESHDEMS-----VAKKDDLOLQOEVLQSES---DOLKEN 1464
 Db 2020 DLEVVOETEKLCLEKONENKQVIVCLEBEELSVMVTSERNQLRGELDTMSKTTALDQSEK 2079
 Qy 1465 IKGIIVAKHLETEEB---LKVAHCCLKEQETINELRVNLSEKETEISTIQKOLEAND 1519
 Db 2080 MKE--KTOELESQSECLHCIOVAEAEVKEKTELTQTLSSDVSSELLKQTHQKLOESLEK 2138
 Qy 1520 -----KLONKIOEIVEKEQNLNIKQISEVOENVNELKQFKEHRKAKDS--ALOSIE 1568
 Db 2139 DSQALSUTKCELEHQIAGL--NKEKELLVKESESLOARLSE---SDYEKLVNSKALEAAL 2193
 Qy 1569 SKMLETNRLQESQEEIQIMIK-----EKEEMKRVQBALQI-----ERDQIKENTK 1614
 Db 2194 VEKGEFALRLSSTQEBVHQLRGIEKLRVIRTEADEKKQLHIAELKERERENDSLDKOVE 2253
 Qy 1615 EIVAKMKEQSEKQVFLKMTAVNETQKMCLEHLKEQFETQ-----KLNLENIETN 1667
 Db 2254 NLERBLOWSEBQ-----ELVILDAENSKADEVETLTKTOEEMARSLKIFELDLVILRSK 2308
 Qy 1668 IRLTQILHEN-----LEEMRSVTKEDR--DLRSVEET---LKVERDOLKENLRET 1712
 Db 2309 ENLTQIQOEKQOLSELKLLSSFKSLLEKEQAQIQIEESKTAVEMLQNLKE--LNEA 2367
 Qy 1713 ITRDLEKQFELKIVMHML-----KSHQ--ETIDKURG--IVSEKTNIEINMQ--KOLHNS 1762
 Db 2368 VAALCGDQIEIMKATQSLDPPTEEBHQLRNRSIEKLRARLEADEKKQLCVLOQLKSEHHA 2427
 Qy 1763 DALKAQDLKIQEBLRIAHMHLKEQOFTIDKLRGIVSEKTDKLSNMQKDLSENAKLOEKI 1822
 Db 2428 DLLKGRVENLELEBIA---RTNOE-----HALEARNSEKGEVETLKAKIEGWT 2473
 Qy 1823 QELKANEHOLIITLKKD---VNNETOKK-----VSEMQLKQIKQDOSL 1861
 Db 2474 QSLRGELDVVTIRSEKENLTLNLOEQRISLEILINSSPENILQEKQEKQVOMKEKSS 2533
 Qy 1862 TLKSL---EIEINL--LAQELHENLEBMKSVMKERNLRRVETLKLRLDOLKESLOETKA 1917
 Db 2534 TAMEMLQTLKELNERNVAALHNDQEAACKA---KEQNLSQVECELEKALQLGLDEAKN 2590
 Qy 1918 RDELETQOELKTARMLSKHEKTVDKLRKIKSEKTIQISDIOKDLDSKD-----1966
 Db 2591 NYIVLQSSVKGIIQEVEDGKLEKDEEISILKQIQOQEQIVSKLSQVEGEHQWLWKEQ 2650
 Qy 1967 -----ELQKIQOELQKELQLLRVKEDVNMSSHKI--NEMEQLKQKQFEPNYLCKCEM 2016
 Db 2651 NLELRNLITVELBQKLOVLOSKNASLODTLEVLOQSSYKNLENELELT------M 2698
 Qy 2017 DNFQITKLIHESLETRIVAKERDELRRIKESIKMERDQFIATIRMIARDRONHQVKPE 2076
 Db 2699 DKMSFVEKVN-----KMTAKETELQREHMA-----OKTAELOE-----ELSGE 2738

770 MHFAELQKFLSQSEHKILHDQHCSSKMSSELQTYVDSLKAEINLVSLNRFQDGLV 829
1075 ENITEMIE-----NOEELLLGDELKQOEIIVAOEKHAI--KKGELSR- 1118
830 KEMQLGLEGLVPSLSSSCVPSDSSSLGDS-SFYRALLETQTDMSLLNLEGAVSANQ 888
1119 -----CDRLAEVEEKLEKKSQQLQKQOQLNVQSEMOMKINEIENLKNELNKE 1171
889 CSUDEVCSSLOENLRTKTPSPAPAKGVELESLECYVRSLEKLEKMSQIGMKNKE 948
1172 L-TLEH-METERLEL-----AQKNENYEVVKSITKERVUKLEKQSFETERDHLRGVIRE 1225
949 IQLEQLSSERQELDCRLQYLSENEQWQOKLT---SVTILEMSKLAEEK-----KQ 998
1226 IEATGLQTKBELKTAHILKHEQETIDELRSV-----SEKTAQIINTQ-DLEKSHTKLOE 1280
999 TEQLSL-----ELEVARIQL-----QGLDLSRSLGIDTEDAIQGNESCDISKHTSETT 1050
1281 EIPVLHE-----EOELLPNVKVKSET----- 1301
1051 ERTFKHDVHOICDKDAQODLNDLIEKITETGALKPTGSCGEGSPDNTNVEPPGBDKTQGS 1110
1302 -----QETMNEBELTEQSTTKOSTTIL----- 1323
1111 SECISELSPGNALVPMDFLGNQEDIHNLQVRKETSNEENLRIHVIEDRDRKVESLLN 1170
1324 -----ARIENRRLINEKFQ-----ESQEEIKSLTKERDNLKTIKEALEVKHDOL----- 1368
1171 EMKELDSKHLQEVOLMTKIBACHELEKIVGELKKNESDLSEKLEFYSCDHOELLQORVET 1230
1369 -----KXHIRETLAKIQESQKOSQOSLNKMEKONETTKIVSEMBOFPKP 1412
1231 SEGINSLEMAHDKSSREDIGDNVAKVNSW--KERFLDV--ENELSKIRSEKASIE-H 1284
1413 DSALRIBIEMGLSKRLQORSHDEMS-----VAEKDDDLQLOEVLQSES-- 1458
1285 EALYLEADLEVQTEKLCLEKONENKQVIVCLEBELSVTTSERNQELGELDTMCKTTA 1344
1459 -DOLKENIKIIVAKHLETEEB-----LKVHCLAKOQETINELRVNLSEKETEISTIQK 1512
1345 LDQSEKMKK-KTOELSHQSELCHCIOVAEAEVKEKTELJQTLSSDVSSELLKDKTHLOE 1403
1513 QLEAIND-----KLONQIBIYEKEBQNLNIKOISEVOENVNELKQPKHEHRKAKDS 1562
1404 KLOSEKDSOALSILFKCELENIQAL-NKEKELLVKSSESLQARLSE--SDYEKLNVS 1458
1563 -ALQSTESKMLETNRLQESQBEIOMIK-----EKHEMKRVQERAIQI-----ERD 1607
1459 KALEAALVEKGBFALRLSSTQBEVHQLRGIEKLVRIEADBKQLHIAEKLKEREREND 1518
1608 QLKENTKEITVAKWKEQKEYQFLKMTAVNETQEKMCETIEHLKQEPETQ-----KML 1660
1519 SLKDKVENLERELQSEENQ-----ELVILDAENSKAEVETLKTQIEEMARSLKVPFELD 1573
1661 ENIETENIRLTQILHEN-----LEEMSVTKERD--DLRSVEET--LKVERDQ 1705
1574 VTLRSEKENLTKIQIEKQQLSELDKLSSFKSLLEEKEQAEIQIKESKTAVERMLQNLQ 1633
1706 KENLRBTITRDLKQBELKIVEMHL-----KEHQ--ETIDKLRG-IVSEKTNWISNQ-- 1755
1634 KE-LNEAVALGQDQIMKATEQSLDPPTEEBEHLRNSIEKLRLEADEKXKQCVIQL 1692
1756 KOLEHNSDALKAQDLKIOELRIFAHMLKEQOBTIDKLRGIVSEKTDKLSNMKOLENSN 1815
1693 KSEEHADLLKGVENLERELETA-----RTNQE-----HALEAENSKGEVETLK 1738
1816 AKLOEKIQIELKANEHOLITLKKD-----VNETQKK-----VSEMQLKK 1854
1739 AKTEGMTQSLRGELDVVITRSEKENLTNELQKEBERISELEITINSFENILOKEQEKV 1798
1855 QIKQDSITLSKL---BIENLN-LAQELHENLEEMKSVMKERDNLRRVEETLKLERDOLKE 1910
1799 QMEKESSTAMEMLQTLKELNERNVAALHNDQEACKA---KEQNLSSQVECLELEKALQLQ 1855

1911 SLOETVARDLEIQOELKTARMLSKHEKETVDKLRKISEKTIQISDIQKDLKSKD--- 1966
1856 GLDEAKNNYIVLOSNGLIQVEDDGKQLEKDEEISRLKNQIQOQOQLVSKLSQVEGE 1915
1967 -----ELOKKIOELOKQELQLLRVKEDVNMHKKI--NEMEQLKQKQEPN 2009
1916 HQWKQENLEIRNLTVLEEQIKIQVQKSNASLOTLVLOSSYKYLENELELT- 1969
2010 YLCKCENDPNQTLTKLHESLEERIVAKERDELRIKESLKMEDQDQIATIREMIARDQ 2069
1970 -----NDKMSFVEKVN-----KMTAKETELQEMHEMA-----OKTAELOE----- 2005
2070 NHQVPEKRLLSQOQHLMESLREKCSRIKELLAKRYSEMDDHYECLARLSLDLKEIIEFH 2129
2006 --ELSGEKNRLAGELQLLEIEIKSSKQDLKELTLENSELKSLDCMKA--DOVKE----- 2057
2130 RIMKKLVLSYVTKIEEQHECINKPMDFIDVEVEKQKELLIKIOHLQDQCVPS-REL 2188
2058 -----QKVREE-----IAEQLR-LHAEEKHQALLDNTKNQYVEIQTYRE- 2098
2189 RDLKLNQMDLHIEIILKDFSESEFPISIKTEFOQVLNKRKEMTQFLEBWLNRDIEK 2248
2099 --KLTSK-----EECLSS-QKLEIDLKSSKEBELNLSLKATQILEELKTKTMDNLKV 2149
2249 NGIOKENDRICQVNNFFNNRIIAMNESTEPEERSATISKEMEQDLKSLKEKNEKLPKNY 2308
2150 NQLAKKENERA-----QGMKLLIKSCQKQLEKEILQKELSO-LQAAQEK----- 2193
2309 QTLKATSLASGAQVNPPTQDNKNPHTSRATQITTEKIRELENSLHEAKESA---MHKBSK 2365
2194 -----QKTGTVMDTK-----VDELTTE-IKELKETLEEKTEADEYLDKYCS 2234
2366 IIRKQKELEVTNDIIAKLOAKVHESNKCLEKTKETIQ--VLQDKVALCAKPYKEIED-L 2422
2235 LLISHEKLEK-----AKEMLETQVAHLCSQKSDRSRGSPLLGVPVGPSPISVTEKRL 2289
2423 KMKLVKIDLEKKNKAKEFEKISATKATVE-YQKEVIRLLARENRSQAQDTSVISHT 2481
2290 SSGQNKASGRQSRSGIWENGSGGPTTAPFESFKSKKAVMSGI---HPAEDT-----EGT 2342
2482 DPQPSNKELTCGGSG-----IVQNTKALILKSEHIRLEK-EISKLQOQNEQLIK 2530
2343 EFEPGLBEVVKGFADIPDTGKTSYIILRTTMTATSPRLAAQKLLSPLSGENLAE 2402
2531 QKNELLSNNQHLNNEVKTWKERTLKREAHKQVTCNSPKSPKVTGASKKKQITPSOCKE 2590
2403 SSKPTAGGR--SQKV-----AQRSPVDSGTILREP--TTKSVPV 2440
2591 RMLQDPVKESPKSCFFDSRSKSLSPHPVRVYFDNSSLGLCPEVQNAGAESVDSPGPWH 2650
2441 NNLPSRPTDSPREGLRVKRGRLVPSP-----KAGLESKGSB----- 2477
2651 ASSGKDVPECKTQ 2663
2478 -----NCKVQ 2482

RESULT 10

US-09-595-684B-23
; Sequence 23, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; TITLE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B

```

; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 1231
; TYPE: PRT
; ORGANISM: Human
US-09-595-684B-23

Query Match
Best Local Similarity 26.3%; Pred. No. 6.4e-37;
Matches 343; Conservative 226; Mismatches 479; Indels 254; Gaps 43;

QY 7 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDFRVFHHGNETTKNVEE 65
Db 10 VRVALRCRPLVPKEISEGQCMCLSFVPGEQVVGTD--KSFTYDFVDFPSTQEVEFNT 67
QY 66 IAAPILDSAIQGYNGTIFAYGQTASGKTYTMG-----SEDLGVIPRAIHDFOKIK 118
Db 68 AVAPILKGVFGYNATVLAQYQSGKTYSMGAYTAQENETVGVIPRVIOQLLFEKD 127
QY 119 KPDPREFLRVSWEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVYVTSMA 178
Db 128 KKSDFEFTLKVSLEYETINEEILDLLCPREKAQINIREDPKEGKIVGLTEKTVLVALDT 187
QY 179 LKWITGKSRHYGETKMRORSRSHITPRMILSREKGEPSNCEGSKVSHNLVLDLAG 238
Db 188 VSCLEQGNRSRTVASTAMNSQSSRAIFTI-----SLEQGGKSKDNSSFR-SKLHLVDLAG 243
QY 239 SRAAQTGAAGVRLKEGNCNINSLFILGOVVKLSDGVGGFINYRDSKLTIRLQNSLGG 298
Db 244 SERQKTKAEGDRLEKGINLRGLCLGNVISALGDDKGGFAPYRDSKLTIRLQNSLGG 303
QY 299 NKTRITICITPV--SFDETLTALQFATAKMNKNTPYNEVSTDEALLKRYRKEIMDLK 356
Db 304 NSHTLMACVSPADSNLEETLRLYADRARKIKKPIVN-----IDPQTAELNHLK 355
QY 357 KOLEEVS-----ETPAQAMEKDQALQLLEEKDLQKQVNEKIENL 397
Db 356 QVQVQLVLLQAHGHTLPGSTVTPSENLQSLMEKNQ--SLVEENEKLSRGLSEAGQT 413
QY 398 TRML---VTSSLTLOELKAKRKRTWC---LGKINMKNSYADQFNIPNTIKTH 451
Db 414 AQMLERLIWTEQANERKNALBELRCHAAKLDLQKLVELQELKENVEIICNLQOLLT 473
QY 452 KLSINLLREIDSVCSESDVFNLTDLSEIWNPAK-----LNQENI 496
Db 474 QLS-----DETVAAMAAIDTAVEQEAQVETSPETSSRSDAFTTQHALRQAQMSKELV 526
QY 497 E-----SELNSLRADY-DNL---VLDYQLRTEKEEMELKLK-EKN 532
Db 527 ELNKALALKEALARKMTQNDSQLQIYQYQNIKEPELEVINLQKBEELVLELQATK 586
QY 533 DLDEFEALERTKQDQEMQLIHEISNLKNLVKRVYNDLENELSSKVLELREKEDQIK 592
Db 587 DANQAKLSERRKRILELE--QIADLKKL-----NEQSKLLKKESTERTVS 633
QY 593 KLOEVIDSQLENIKMDLSYSLESDPKOMQTLFPAETVAL---DAKRESAFIRSE-N 648
Db 634 KLNQEIWMKMNQVQL-MRQMKDEAKFRQWQKR-DKEVIQLKERDRKQVELLKLERN 691
QY 649 LS-----LKEKMKELATTYKQEMENDIQLYSQLEAKKQVQDLEKLOSAFNEIKLSL 703
Db 692 FQKQSNVLRKTEEAAAANKRUDALQ-----KQREVAOKRKEOTOSRGEGTA---- 739
QY 704 IDGKVPKOLLNLELEGKITDLQKELNKEVEEALREEVILLSELK-SLPEVERLRKE 762
Db 740 --ARVKNWLGNEIEVWVSTEEAKRHLNDLEDRKILAQDVAQVQKESGENPPKLRR 797
QY 763 IQDKSEELHIITSEKOKLFSEVHVHESRVQGLLEIGIKTKDLDLATQSNYKSTQDEQNF 822
;

108 TFSLTEVRGOVSESDSITTKQIESLETETEMFRSAQIADLQKLLDASEDR-PKQRWENI 856
109 KTLHMFPEQYKVMVLENERMNOBIVNLSKQAFDSSLGALKTELSTYKTOELQEKTR-- 880
110 KYLIGELVSSKIQVSKLESSKQSKTSKSCADMQMLFEERNHF 908
111 EVOERLNEQKLEQENRDSPLQVVEREKTILITEKLOTLFEVKTILTEKDDLK 935
112 AEIETELQAEIVRNEQ-----QHEKVLVLLSQSQSMAE-----K 945
113 QLOESLQIERDQKSDIHDVTNNMNIQTQEURNALESKQHOETINTLKSKIS--EVSVR 993
114 QLESVSEKQQLLS---TLKQDELEKQREVCEQOQLLRENEIILKQKLTLLQVSR 1001
115 GETKDEFOQKMGVI--DKKQDLFAKNTQTLTADVKD 1035
116 QKHLPKDPTLLSPDSSEFYQPKPKSRVKEKFLQEQMDIEDLKYSCHSEHSNEHEDGDD 1061
117 NEIIEQOKRIFSLI-----QEKNE--LQOMLESVIAKEQL 1069
118 DEGDDEWPTKLVNVRNKNIOCSCKGKQCGCRKQKQSGVDCDDPTKCRNRQ 1121
119 KTDKENIEMTIENQBELRLGLDELKQOEI-----VAQEKHAIKKEGELSRCTD 1120
120 GKSGLGTVERTQDSSESK-----LEDPTVTPGLSFPNVPVCAFPNSKILKE----- 1168
121 RLAEVBEKLEKESQ-----QLEKQOQLLVQEMSEMOKK 1156
122 MCDVEQVLSEKTPPAPSPFDLPKLVHATEYQENKAPGKK 1209

RESULT 11
US-09-592-054-2
; Sequence 2, Application US/09592054
; Patent No. 6440684
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Finer, Jeffrey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6440684e1 motor proteins and methods for
; FILE REFERENCE: 1016
; CURRENT APPLICATION NUMBER: US/09/592,054
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1232
; TYPE: PRT
; ORGANISM: Human
US-09-592-054-2

Query Match
Best Local Similarity 7.1%; Score 941.5; DB 4; Length 1232;
Matches 341; Conservative 227; Mismatches 480; Indels 255; Gaps 43;

QY 7 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDFRVFHHGNETTKNVEE 65
Db 10 VRVALRCRPLVPKEISEGQCMCLSFVPGEQVVGTD--KSFTYDFVDFPSTQEVEFNT 67
QY 66 IAAPILDSAIQGYNGTIFAYGQTASGKTYTMG-----SEDLGVIPRAIHDFOKIK 118
Db 68 AVAPILKGVFGYNATVLAQYQSGKTYSMGAYTAQENETVGVIPRVIOQLLFEKD 127
QY 119 KPDPREFLRVSWEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVYVTSMA 178
Db 128 KKSDFEFTLKVSLEYETINEEILDLLCPREKAQINIREDPKEGKIVGLTEKTVLVALDT 187
QY 179 LKWITGKSRHYGETKMRORSRSHITPRMILSREKGEPSNCEGSKVSHNLVLDLAG 238
Db 188 VSCLEQGNRSRTVASTAMNSQSSRAIFTI-----SLEQGGKSKDNSSFR-SKLHLVDLAG 243

```

RESULT 12
US-09-592-054-8
; Sequence 8, Application US/09592054
; Patent No. 6440684
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Finer, Jeffrey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6440684el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1016
; CURRENT APPLICATION NUMBER: US/09/592,054
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1234
; TYPE: PRT
; ORGANISM: Human
; US-09-592-054-8

Query Match 6.9%; Score 924.5; DB 4; Length 1234;
Best Local Similarity 25.8%; Pred. No. 6.6e-36;
Matches 341; Conservative 228; Mismatches 398; Indels 357; Gaps 49;

QY 7 YACVVRPLNRSREESLG-ETAYQVTKTDNNVIYQVDSKSPNDRVPHGNETTKNVYEE 65
DB 10 VRVALRCRPLVPKEISEGCMCLSFVPGTQVVVGTD--KSTYDFVDFDPTCQEBVENK 67
QY 66 IAAPIIDSAIGYNGTIFAYGTAGTKTYTMG-----SEDLGVIPRAIHDFQKIK 118
DB 68 AVAPLLKIGIFGYNATVLAAYQGTSGTKTYSNGGATYAEQENETVGIIPVQLLFKEID 127
QY 119 KFPDREFLLRYSYMEIYNETITDLCGQPKMKPIIREDVNRNVVADLTREVVVTSEMA 178
DB 128 QKSDFEFTLVKSYLEIYNEEILDLCPREKAQINIRDPKEGIVGTEKTVLVALDT 187
QY 179 LKWTTKGSKSHYGETQKQSSSHITFRMLLESREKGEPSNCEGSKVSHNLNVDLAG 238
DB 188 VSCLEQGNNSRTVASTAMNSQSSRSHAFIT---SLEQKKSKDKNSFR-SKLHLVDLAG 243
QY 239 SERAAQTGAAGVRLKEGNCINRSFILGVTKKISDGOVGFYINRDSKLTILQNSLGG 298
DB 244 SERQKTKAEGDLKEGINIRGLCLGNVISALGDDKGSFVPYRDSKLTLLQDSLGG 303
QY 299 NPKTRIICTIPV--SFDETLTALQFASATKYMNTPVYNEVSTDEALLKRYRKEIMDLK 356
DB 304 NSHTLMIACVSPADSNLEETLSTRYADRARKIKNPIVN-----IDPHTAELNHLK 355
QY 357 KQLEEVSL-----ETRAQAMEKQDLAQLLEKD---LLOKVQNEK 393
DB 356 QOVQOLQVLLLOAHGGTLPGSINAPSNLOSLMEKNQ--SLVEENEKLSRCLSKAAGT 413
QY 394 IENLFRMLVTSSSLTLOQELKAKRK--RRVTCWLGKINK---MKNSNYADQFNPTNT 447
DB 414 AQMLERIILTEQ---VNEKLNKLESLRQHAACKLDLQKLVELEDQELKENVEIICNLQ 470
QY 448 TKTHKLSINLREIDSVCSDESDFNTLDTLSEIENWPNATKILMOENIESELNRLRY 507
DB 471 QLIITQLS-----DETVAATAAIDTAVEEAQVETSPET-----SRSSDA 510
QY 508 DNLVIDYEQLRTEKEEMELKKEKNDLDEFFALERK-TKKQOBMOJI--HEISNLKNLVK 564
DB 511 FTTQHALHOAQMSKEVEL---NNALAKLEALVRQOTQNDQLOLQIQFYQYQDNINKNL-- 564
QY 565 HREYVNOQLENELSSKVELLREKEDQIKKLQYIDSQKLENIKMDLSYSLSEIDPKQMK 624
DB 565 -----ELE-----VINQKEKEELVRELQ----- 583
QY 625 QTLFDAETVALDAKRESAFIRSENLEKEMKELATYKQMDENDIOLYQSOLEAKKMQOV 684
DB 584 -----TAKKNQAQAKLSEHR-----RKLQJ- 603

QY 239 SERAAQTGAAGVRLKEGNCINRSFILGVTKKISDGOVGFYINRDSKLTILQNSLGG 298
DB 244 SERQKTKAEGDLKEGINIRGLCLGNVISALGDDKGSFVPYRDSKLTLLQDSLGG 303
QY 299 NPKTRIICTIPV--SFDETLTALQFASATKYMNTPVYNEVSTDEALLKRYRKEIMDLK 356
DB 304 NSHTLMIACVSPADSNLEETLSTRYADRARKIKNPIVN-----IDPHTAELNHLK 355
QY 357 KQLEEVSL-----ETRAQAMEKQDLAQLLEKD---LLOKVQNEK 393
DB 356 QOVQOLQVLLLOAHGGTLPGSINAPSNLOSLMEKNQ--SLVEENEKLSRCLSKAAGT 413
QY 394 IENLFRMLVTSSSLTLOQELKAKRK--RRVTCWLGKINK---MKNSNYADQFNPTNT 447
DB 414 AQMLERIILTEQ---VNEKLNKLESLRQHAACKLDLQKLVELEDQELKENVEIICNLQ 470
QY 448 TKTHKLSINLREIDSVCSDESDFNTLDTLSEIENWPNATKILMOENIESELNRLRY 507
DB 471 QLIITQLS-----DETVAATAAIDTAVEEAQVETSPET-----SRSSDA 510
QY 508 DNLVIDYEQLRTEKEEMELKKEKNDLDEFFALERK-TKKQOBMOJI--HEISNLKNLVK 564
DB 511 FTTQHALHOAQMSKEVEL---NNALAKLEALVRQOTQNDQLOLQIQFYQYQDNINKNL-- 564
QY 565 HREYVNOQLENELSSKVELLREKEDQIKKLQYIDSQKLENIKMDLSYSLSEIDPKQMK 624
DB 565 -----ELE-----VINQKEKEELVRELQ----- 583
QY 625 QTLFDAETVALDAKRESAFIRSENLEKEMKELATYKQMDENDIOLYQSOLEAKKMQOV 684
DB 584 -----TAKKNQAQAKLSEHR-----RKLQJ- 603

```

QY 685 DLEKELQAFNEITKLTSLIDGKVPKDLLCNLELGGKITDLOKELNKE-----VENEALRE 741
Db 604 -----ELEGQIALDKKLNQSKLLKKESTER 631
QY 742 EVILLS-ELKSLPSEVERLKEIQDKSEELHITSEKDKLFSEVHVHKSVOGLLEIGK 800
Db 632 TVSKLANQIRMMKNQVOLMRQKDAEFQWQKQKDK---EVIQLKDRKQVYELLK 688
QY 801 TKDOLATTQSNYKSTQDFQNEKTLHMDPEQYKMWLEENRM-----NQEIUNLSKEA 854
Db 689 LE-----RNFQKSNVLRKKEEAAAANKRLDALQKQREVADKRRKET 731
QY 855 QK--FSSSLGALKTELSTQELQEKTRVQOERINEM-----EOLKEQLENRDS 901
Db 732 QSRGMEGTAARVRNWLGNEL-EVMVSTEEAKRHLNLDLDRKILAQDVVLKKEKESREN 790
QY 902 PLQTVREKTLITEKLOOTLEEVKTLTQEKDDLKQIQESQIQRDQOLKSDIHDVTANMID 961
Db 791 PPPKLRKCTFSLSEVHGQVLESDCIT-----KQI-ESLETEMELRSAQIADLQOKLLD 843
QY 962 TOBOLNALESKHOHRTINT-----LKSKESEVSRNLMHEENTGETKDFQOKWVG 1014
Db 844 AESEDRP-----KQCHENTATILEAKCAUKYLIGELVSSKIHV----- 881
QY 1015 IPKQDLEAKNTQTLTADVDKNEIIBQQRKIFSLIQEKNELOQMLSEVIAEKEQLKTDLK 1074
Db 882 --TKLENSLRQSKASCADM-----QKMLF--EEQNH-----SEIETEIQ 917
QY 1075 -ENIEMTINOELRLILGDLBKQOQIIVAQEKHNAIKKGEFSRTCDRLAEVBEKLEKS 1133
Db 918 ABLVRMEQHQEKVLYLVLSQESQMAEKQLEKASKEPQLVST-----LQCODELEKM 973
QY 1134 QOLQEQOQQLNVQOEMSEMOKKINIEINLKNELKNKELTLE-----HMETERLEAOK 1187
Db 974 REVCEQOQQL--QE-----NEI-----IKOKLILLOVASRQKHLPNDTL--LS 1013
QY 1188 LNEYEEVKSITKERKVL--KELOKSPETERDHLRGVIRIEATGLQTKELKIAHIH-LK 1245
Db 1014 PDSSPEYIPPKPSRVKEKFLQSDMI-----EDLKVCSEHSVN 1053
QY 1246 EQHE 1249
Db 1054 EHED 1057

```

RESULT 13

```

US-09-722-139-2
; Sequence 2, Application US/09722139
; Patent No. 6355471
; GENERAL INFORMATION:
; APPLICANT: Berard, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No.6355471el motor proteins and methods for
; FILE OF INVENTION: their use
; FILE REFERENCE: 1055
; CURRENT APPLICATION NUMBER: US/09/722,139
; CURRENT FILING DATE: 2000-11-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: Human
US-09-722-139-2

```

```

Query Match          6.7%; Score 889; DB 4; Length 1375;
Best Local Similarity 26.2%; Pred. No. 3.4e-34;
Matches 380; Conservative 244; Mismatches 537; Indels 292; Gaps 56;

```

```

QY 6 AVAVCVRRPLNSRESLGETAQVWTKTDNNVIYQV-----DGSKSFNDRVF 53
Db 3 SVKAVVRPMRNRREKLEAKFIIQMEKSKTTITNLKIPEGTGDSGRERTKTFYDFSF 62

```

```

QY 54 HGNET-----TKNYVEEIAAPIIDSAIQYNGTIPAYCQTASGKTYTMGSDHDLGVI 106
Db 63 YSADTKSPDVSQEMVFKTLGTDTVKSAPFEGNACVPAYCQTGSGKSYTMGNSGDSGLI 122
QY 107 PRAIHIDIFOKIK--KFPDREFLLRVSYMEIYNETITDLL-CGTQKMKPLIREDVNRNV 162
Db 123 PRICEGLFSEINETTWDEASFTFEVSYLEIYNERVDDLRRKSKSTFNLRVREHPKEGP 182
QY 163 YVADLTVEVVYTSSEMALKWITKGEKSRHYGETKQNRSSRSHTIFRMILESREKGEPSNC 222
Db 183 YVEDLSKHLVQYGDVEELMDAGNINRTTAATGMDVSSKSHAFITTKFTQAKFDSMPCC 242
QY 223 EGSVKYSHLNLVDLAGSRAAQTGAAGVRLKEGCNINRSLFILQGVIKKLSDOQVGGFIN 282
Db 243 E---TVSKIHLVDLAGSERADATGATGVRLKEGGINKSLVTLGAKKKQV-----FVP 292
QY 283 YRDSKLTIRLONSLGNPKTRIICTITP--VSPDETLTALOPASTAKYMKNTPVVNEVST 340
Db 293 YRDSVLTWLLKDSLGNSKTIATISPADVNYGETLSTLRYANRANKIINKETINE-DA 351
QY 341 DEALLKRYKEIMDKKQLEEVSLSTRAQAMEKDLAQL-----LEEKDLLEKQVNE- 392
Db 352 NVKLIRELRAETARLKTLL-----AQNQIALLDSPTALSMEEK--LQ--QNEA 396
QY 393 KIENLTREMLVTSSLT---LQOELKAKRRRVTCWLGK-----INK 430
Db 397 RVOELTEKWNKWNKTONILKEQTTLALRKEGIVGLDSELPHLIGIDDDLLSTGIYHL 456
QY 431 MKNSYADQFNPTNITTTKHLKSLINLLREIDSVCSDESDFSTNTLDTLSEIEMNPATKL 490
Db 457 KEGQTVVGDDASTEQDIVLHGLDLE-----SEHCIFENIGTGTVTLPLSGSQCS 506
QY 491 LNOEMIESELNSLRADYDNLVDYEQLRTEKEMELKJKEKNDLDBEALERTKTKDQEM 550
Db 507 VNGVQI-VEATHLNOQAVILLGRTNMFNFHPKEAKLREK-----RKS 549
QY 551 QLIHIEIS-NLKNLVKRE-----VYNQDLENELSKVELLREKEDQIKKLOVYDSOKL 603
Db 550 GLSSFSLSMTDLSKSRNLSAVMLYNPGLFEFQOREEL--EKLESKRKLIEEME-EKQ 606
QY 604 ENIKMDLSYSLESIEDPKOMKOTLFDATVALDAKRESAFRLSENLELEKEMKELATTYK 663
Db 607 KSDKAELERMQQEVZ--TQKET-----EIVQIQIRQESLAKRSRPHENKLDLAEKE 660
QY 664 QMENDIQLYSQOLEAKKQMDV-----LEKEIQ-----SAFNEITKLTSLI 704
Db 661 KFEERLRQEQEIELQKQREBEETFLRQOELQRLKELNNEKAQKQFQFQELDQKQEK 720
QY 705 DGKVPKDLLCNLELGGKITDLOKELNKEVEENEALREEVILLSELKSLPSEVERLKEIQ 764
Db 721 DEQYAK-----LELE-----KXRLREQE--KEQVML-----VAHLEEQLR 753
QY 765 DKSEELHIITSEKDKLFSFV--VHKESR-VQGLLEEIGTKDDLATQTSNYKSTQDFQFN 821
Db 754 EKQEMIQLLRR-----GEVQWVEEKDLEGIRSELVRKARA-----GGDEGDGE 800
QY 822 FKTLMDF-----EQYKQVLEENRMNQEIUNLSKEAQKFPDSSLGALKTELSTYKTELQ 876
Db 801 LEKAQLRFFEFKRRQLVKLVNLEKDLVQOKDI-LKKEVQEEQEILECLKCEHD-KESRL 858
QY 877 EKTREVOERLNEMEQLEQLENRDSPLQTVREKTLITEKLOOTLE-EVKVTLTQKDDLK 935
Db 859 EKHDESVTDTVEVPQDFEFKIPVEYRLQYKXERQ-----LQYLLQNHLPTLLEEK- 907
QY 936 QLOESLQI-ERDQLKSDIHDVTNMMIDTQEOALNALSBLKQHOETINTL-KSKI SEEVS 993
Db 908 --QRAFELDRGFLSLD-----NTLYQVEKEWEEKEBQLAQTQANANQKQLQAFEFETA 960
QY 994 NLHMEENTGETKD-----EFQKQMVGIDKKQDLKAKNTQTLTADVDKNEIIEQQRKIF 1046
Db 961 NIARQEEKVRKKEKILEESREKQOREALERALANLERRHSALQRHSTLGTLEBQOKLA 1020

```


QY 1047 SLIQEKNELOQMLSVIAEKEQKLTDLKENTIENTIENOEELRLGLDELKQOEIVAEQKN 1106
 Db 1021 SLNQSREQSGIQASLEAEQALEKD-QERLEYEIQLKQ-----KIYEDGVQKD 1070
 QY 1107 HAIKKEGELSRCDRLAEAEVEKLEKESQOIQEQQOQLLNQVEEMSEMOKKINEIENLKN 1166
 Db 1071 H-----HGTL-----EGKVASSSLPVSAAKSHLVPL-----MDARINAY--IEEE 1108
 QY 1167 LKNEKLTLEHMETERLELAQKLNENVEVKSITKERVKLEKQSFETERDHLRGYIREI 1226
 Db 1109 VORRLQDLHRVISEGCGTSADTMKNEKLNHTGIORKL-----KYLCDRLCLVLMPEP 1162
 QY 1227 EATGLQKEELKIAHILKEHQETIDELRRSVSEKTAQIINTQDLKESHTKLOEIPVLH 1286
 Db 1163 DAAACANHPLOQLVQLSLDWKT--EIPDLVLPNGVQVSS-----KFQTTLVDMVIFLH 1215
 QY 1287 BEQELLPNVKVSQTQMTMELELLTQGSTTKDSTTLARIEMERLRLNEKFEQSEBEIKS 1346
 Db 1216 GNMEV--NVPSLAEVQ-----LLLYTTVKVMGDS-----GHDQOCQS 1249
 QY 1347 LTKERDNLKTIKE 1359
 Db 1250 LVLNTHALVKE 1262
 RESULT 14
 US-09-721-832-2
 ; Sequence 2, Application US/09721832
 ; Patent No. 6399346
 ; GENERAL INFORMATION:
 ; APPLICANT: Beraud, Christophe
 ; APPLICANT: Freedman, Richard
 ; TITLE OF INVENTION: No. 6399346el motor proteins and methods for
 ; TITLE OF INVENTION: their use
 ; FILE REFERENCE: 1055
 ; CURRENT APPLICATION NUMBER: US/09/721.832
 ; CURRENT FILING DATE: 2000-11-24
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 1375
 ; TYPE: PRT
 ; ORGANISM: Human
 US-09-721-832-2
 Query Match 6.7%; Score 889; DB 4; Length 1375;
 Best Local Similarity 26.2%; Pred. No. 3.4e-34;
 Matches 380; Conservative 244; Mismatches 537; Indels 292; Gaps 56;
 QY 6 AVAVCVRRPLNSREESLGETAQVYWKTDNNVIYQV-----DGSKSFNPDVRF 53
 Db 3 SVKAVRVRPMNRREKDLKAEFLIQEKSCTITNLKIPGEGTGDGSGRETKTFYDPSF 62
 QY 54 HGNET-----TKNVEIEAIPIDSAIQYNGTIFAYGQTASGKTYTMGSEDLGLVI 106
 Db 63 YSADTKSPDYVSQEMVFKTLGTDVVKVSAPEGINACVAYGQTCGSKSYTMGSGDGLI 122
 QY 107 PRAIHDIQKIK--KFPDRELLRVSVMEIYNETITDLL-CGTQMKPLIIREDVNRV 162
 Db 123 PRICEGLSRINETTRWDASRTEVSYLEIINERVDLRKSKTFNLRVREHPKEGP 182
 QY 163 YVADLTETEEVVTSEMAKWIITKGEKSRHYGETKMNORSRSHITIFMILESREKGPSPNC 222
 Db 183 YVEDLSKHLVQYGVDEELMDAGNINRTTAATGMNDVSRSHAFITKFTQAKFDSEMP 242
 QY 223 EGSVKVSHUNLVDLGASERAQGTAGVTLKEGCNINRSIFILGVVTKLSDGVQGVFIN 282
 Db 243 E---TVSKIHLVDLAGSERADATGATGVRLXEGGNINKSLVLTGAKKKQV-----FVP 292
 QY 283 YRDSKLTILONSLGNGPNTRICTITP--VSFDELTALQFASHTAKYMKNTIPYVNEVST 340
 Db 293 YRDSVLTLLKSLGNGSKTIMIATISPADVNYGETILSTLYANRAKNIINKPTINE-DA 351

QY 341 DEALLKRYKEIMDLKKQLEEVLETPAQAMEKQOLAQL-----LEEKDLLQKVONE- 392
 Db 352 NVKLIIRELRAEIRARLTKLL-----AQGNQIALLDSPTALSWEK--LQ--ONEA 396
 QY 393 KIENLRYMLVTSSSLT---LQOBLKAKRRRVWTCLGK-----INK 430
 Db 397 RVQELIKWTKWNETQNLKEQTALARKGIGVLDSELPHLIGIDDDLLSTGILYHL 456
 QY 431 MKNSNYADQFNIPNITTKTKHSINLRLREIDSVCSSEDFVSTLDTLSEIENWPNATKL 490
 Db 457 KEGQTYVGRDDASTEQDVLVHGLDLE-----SEHCIFENIGTIVTILPLSGQCS 506
 QY 491 LNQENISELNSLRADYDNLVLYEQLETEKESMELKKEKNDLDEFEALERKTKKQDEM 550
 Db 507 VNGVQI--VEATHLNQGAIVILLGRTNMFRFNHPKRAAKLREK-----RKS 549
 QY 551 QLIHEIS-NLKNLVKRE-----VYNQDLENELSSKVELLREKEDQIKKIQEYIDSOQL 603
 Db 550 GLISSFSLMTDLSKSRNLSAVMLYNPGLFEFEQREEL--EKLESKRKLIEEME-EKQ 606
 QY 604 ENIKMDLSYLSIESIEDPKQMKQTLFDAETVALDAKRSAFARSENLELKEKKELATYTK 663
 Db 607 KSDKAELERMQOEVE--TORKET-----EIVOLQIRKQFESLKRKSFHLENKDKOLLAKE 660
 QY 664 QMENDIQLYQSOLEAKKKQVVD-----LEKELQ-----SAFNEITKLTSLI 704
 Db 661 KFEERLEEQOEIELOKRRQEEETFLRVOEELQBLKELNNEKAEKQIFQELQLOKEK 720
 QY 705 DGKVPKOLLNLELEKGTIDLOKELNKEVEENALREEVILLSELKSLPSVERLRKBIQ 764
 Db 721 DEQYAK-----LELE-----KKRLEQOE--KEQVML-----VAHLEEQUR 753
 QY 765 DKSELHIITSEKDKLFSFV--VKESR--VOGLLEBEGTKDLDLATTSYKSTDOEFQN 821
 Db 754 EKQEMIQLRR-----GEVQWVEEKDLEIGIESLRVKEARA-----GGDEGEE 800
 QY 822 FKTLHMF-----EQKVMVLEENRMNOEIVNLSKEACKDSSLGALKTELSTYKTELQ 876
 Db 801 LEKAQLRFEFFPKRRQLVKLVNLEKDLVQOKDI-LKKEVQEEQEILECLKCEHD-KESRL 858
 QY 877 EKTREVBQRLNEMEQLEENRDSPLQTVREKTLITKELQOQTL-EVKTLTQEKDILK 935
 Db 859 EKHDSEVTDVTEVPQDFEKIPVEYRLQYKERQ-----LQYLIQNHLPITLEEK- 907
 QY 936 QLQESLQI-ERDQLKSDIHDVTNNMIDTQOELRNALLESIKQHQETINTL--KSKISEEYR 993
 Db 908 --QRAFEILDRGPLSLD-----NTLYQVEKEMEEKEBQLAQYQANANOLKQLQATFEPTA 960
 QY 994 NLHMEENTGETKD-----EFQOQWVGIDKKQDLKAKNTQTLTADVKDNEIIEQOKRIF 1046
 Db 961 NIARQEEKVRKKEKILESREREKQREALERALARLERRHLSALORHSTLTGTEIEEQOKLA 1020
 QY 1047 SLIQEKNELOQMLSVIAEKEQKLTDLKENTIENTIENOEELRLGLDELKQOEIVAEQKN 1106
 Db 1021 SLNQSREQSGIQASLEAEQALEKD-QERLEYEIQLKQ-----KIYEDGVQKD 1070
 QY 1107 HAIKKEGELSRCDRLAEAEVEKLEKESQOIQEQQOQLLNQVEEMSEMOKKINEIENLKN 1166
 Db 1071 H-----HGTL-----EGKVASSSLPVSAAKSHLVPL-----MDARINAY--IEEE 1108
 QY 1167 LKNEKLTLEHMETERLELAQKLNENVEVKSITKERVKLEKQSFETERDHLRGYIREI 1226
 Db 1109 VORRLQDLHRVISEGCGTSADTMKNEKLNHTGIORKL-----KYLCDRLCLVLMPEP 1162
 QY 1227 EATGLQKEELKIAHILKEHQETIDELRRSVSEKTAQIINTQDLKESHTKLOEIPVLH 1286
 Db 1163 DAAACANHPLOQLVQLSLDWKT--EIPDLVLPNGVQVSS-----KFQTTLVDMVIFLH 1215
 QY 1287 BEQELLPNVKVSQTQMTMELELLTQGSTTKDSTTLARIEMERLRLNEKFEQSEBEIKS 1346
 Db 1216 GNMEV--NVPSLAEVQ-----LLLYTTVKVMGDS-----GHDQOCQS 1249
 QY 1347 LTKERDNLKTIKE 1359

[illegible]

Search completed: July 29, 2004, 09:42:43
Job time : 77.178 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:28:21 ; Search time 39.7887 Seconds
(without alignments)
6437.961 Million cell updates/sec

Title: US-10-045-631B-88
Perfect score: 13329
Sequence: 1 MAEGAVAVCVVRPLNSRE.....SOPGPHASSGKDVPCKTQ 2663

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13329	100.0	2663	1 S28261	centromere protein
2	3615.5	27.1	2954	1 T14156	kinesin-related pr
3	1127.5	8.5	3259	1 A55539	giantin - human
4	1125.5	8.4	3225	2 I52300	giantin - human
5	1089	8.2	3187	2 J55837	364K Golgi complex
6	1056	7.9	1909	2 A45592	liver stage antigen
7	1048.5	7.9	1388	2 T30335	KLP2 protein - Afr
8	1026	7.7	2442	2 T08621	centrosome associa
9	1004	7.5	823	2 T52425	kinesin-like prote
10	980.5	7.4	1780	2 T17272	hypothetical prote
11	971.5	7.3	888	2 D56619	microtubule-associ
12	952.5	7.1	1231	2 A45803	repeat organellar
13	946	7.1	1939	2 T18372	kinesin-like prote
14	935.5	7.0	1226	2 I51617	kinesin-like prote
15	929.5	7.0	1742	2 T49451	kinesin-like prote
16	917.5	6.9	1225	2 A56514	chromokinesin - ch
17	900	6.8	2253	2 T30336	nuclear/mitotic ap
18	898	6.7	1459	2 T30196	kinesin motor prot
19	896	6.7	1738	2 T14867	interaptin - slime
20	893.5	6.7	1957	2 T38077	hypothetical colle
21	891.5	6.7	1875	2 S38173	myosin-like protei
22	890	6.7	1979	2 C71622	hypothetical prote
23	887.5	6.7	1229	2 T48959	kinesin-like prote
24	880.5	6.6	2829	2 A42771	reticulocyte-bindi
25	878	6.6	2139	2 T18296	myosin heavy chain
26	873.5	6.6	1827	2 T16270	hypothetical prote
27	867.5	6.5	963	1 A41919	kinesin heavy chai
28	867	6.5	1898	1 A54973	trichohyalin - hum
29	860	6.5	2331	2 T25410	hypothetical prote

30	857.5	6.4	2094	2 S33124	tpr protein - huma
31	856	6.4	967	1 A35075	kinesin heavy chai
32	854.5	6.4	2401	2 T28676	thoptry protein -
33	854	6.4	1805	1 A64224	hypothetical prote
34	848.5	6.4	975	1 A31497	kinesin heavy chai
35	846	6.3	1121	2 T06065	hypothetical prote
36	844	6.3	1031	1 A38713	kinesin heavy chai
37	843	6.3	2269	2 T28677	thoptry protein -
38	839.5	6.3	2166	2 G70163	hypothetical prote
39	835.5	6.3	1032	2 I38510	neural kinesin h
40	831	6.2	1058	2 T47525	kinesin-related pr
41	830	6.2	1027	2 S37711	kinesin heavy chai
42	830	6.2	1066	1 A48669	kinesin-related pr
43	828	6.2	1837	2 T41023	probable nuclear p
44	828	6.2	3885	1 A21605	dystrophin, muscle
45	826.5	6.2	1263	2 T13465	hypothetical prote

ALIGNMENTS

RESULT 1

S28261
centromere protein E - human
N:Alternate names: centromere 312K protein; kinesin-related protein CENP-E
C:Species: Homo sapiens (man)
C>Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C:Accession: S28261
R:Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
A:Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.
A:Reference number: S28261; MUID:93024922; PMID:1406971
A:Accession: S28261
A:Molecule type: mRNA
A:Residues: 1-2663 <YEN>
A:Cross-references: EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865
C:Genetics:
A:Gene: GDB:CENPE
A:Cross-references: GDB:361164; OMIM:117143
A:Map position: 4q24-q25
C:Superfamily: centromere protein E; kinesin motor domain homology
C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
F:7-335/Domain: kinesin motor domain homology <KNOT>
F:86-93/Region: nucleotide-binding motif A (P-loop)
F:486-2183/Domain: coiled coil #status predicted <COI>
F:92/Binding site: ATP (Lys) #status predicted

Query Match		100.0%;	Score 13329;	DB 1;	Length 2663;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 2663;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MAEGAVAVCVVRPLNSRSLG	TAQYVKTNNVYQVDGSKSFN	DRVFHGNETTK	60
Db	1	MAEGAVAVCVVRPLNSRSLG	TAQYVKTNNVYQVDGSKSFN	DRVFHGNETTK	60
QY	61	NYVEIAAPIIDSAIQGYNGT	IFAYGQTASGKTYTMGSEDL	GLGVI PRAIHDFQIKKFF	120
Db	61	NYVEIAAPIIDSAIQGYNGT	IFAYGQTASGKTYTMGSEDL	GLGVI PRAIHDFQIKKFF	120
QY	121	PDREFLLRVSYWEIYNETIT	DLCCGTQMKPLIITREDVNR	VNVYVADLTVEEVVYVTS	EMALK 180
Db	121	PDREFLLRVSYWEIYNETIT	DLCCGTQMKPLIITREDVNR	VNVYVADLTVEEVVYVTS	EMALK 180
QY	181	WITKGEKSRHYGETKMNQ	RSRSHITFRMILESREKGE	PSNCEGSKVSHLNLVDL	AGSE 240
Db	181	WITKGEKSRHYGETKMNQ	RSRSHITFRMILESREKGE	PSNCEGSKVSHLNLVDL	AGSE 240
QY	241	RAAQTAGVRLKEGCNINR	SLFILGVIKLSQGVGGFIN	YRDSKLTIRLQNSLGNP	300
Db	241	RAAQTAGVRLKEGCNINR	SLFILGVIKLSQGVGGFIN	YRDSKLTIRLQNSLGNP	300
QY	301	KTRICTITPVSPFETLTAL	QFASTAKYMKNTPYVNE	STDEALLKRYRKEIMDL	KKOLE 360
Db	301	KTRICTITPVSPFETLTAL	QFASTAKYMKNTPYVNE	STDEALLKRYRKEIMDL	KKOLE 360

301 KTRIICTITPVSFDETITALQFASATKYMNTPYVNEVSTDEALLKRYRKEIMDLKKQLE 360 Db

361 EVSLETRAQAWKQDLAQLEEDKLLQKVONEKIENITRLMTVSSSLTLOELKAKKRR 420 Qy

361 EVSLETRAQAWKQDLAQLEEDKLLQKVONEKIENITRLMTVSSSLTLOELKAKKRR 420 Db

421 VTWCLGKINMKNYSYADQFNIPITNITTKHLSINLLREIDESVCSSESDFVFSNTLDTLS 480 Qy

421 VTWCLGKINMKNYSYADQFNIPITNITTKHLSINLLREIDESVCSSESDFVFSNTLDTLS 480 Db

481 EIEWNPATKLINOINIESELNSLRADYDNLVDYEQURTEKEEMELKKEKNDLDEFEAL 540 Qy

481 EIEWNPATKLINOINIESELNSLRADYDNLVDYEQURTEKEEMELKKEKNDLDEFEAL 540 Db

541 ERKTKDOEWMLIHEISNLKNLVKRVYQDNLLENLSKVELLREKEDQIKKLOEYIDS 600 Qy

541 ERKTKDOEWMLIHEISNLKNLVKRVYQDNLLENLSKVELLREKEDQIKKLOEYIDS 600 Db

601 QKLENIKMDLSYSLESTEDPKMQKOTLFDABTVALDAKRESAFRSENLELKEKMKELAT 660 Qy

601 QKLENIKMDLSYSLESTEDPKMQKOTLFDABTVALDAKRESAFRSENLELKEKMKELAT 660 Db

661 TYKQWENDIQYQOLAKKQWVDLEKELQSAFNEITKLSLDGKVPKDLLCNLELEG 720 Qy

661 TYKQWENDIQYQOLAKKQWVDLEKELQSAFNEITKLSLDGKVPKDLLCNLELEG 720 Db

721 KITDQKELNEKEVENEALREEVILLGELKSLPSEVERLRKEIQKSELSHIIITSEKDKL 780 Qy

721 KITDQKELNEKEVENEALREEVILLGELKSLPSEVERLRKEIQKSELSHIIITSEKDKL 780 Db

781 FSEVVHKESTRVQGLEBIGTKDLDLATTQSNYKSTDOBFQNFKTLHMDPEQKYMWLEEN 840 Qy

781 FSEVVHKESTRVQGLEBIGTKDLDLATTQSNYKSTDOBFQNFKTLHMDPEQKYMWLEEN 840 Db

841 ERMQEIVNLSKBAQKFDSSIGALKTELSTYQTOELOKTRVQBELNEMEQKOLENRD 900 Qy

841 ERMQEIVNLSKBAQKFDSSIGALKTELSTYQTOELOKTRVQBELNEMEQKOLENRD 900 Db

901 SPLQTVREKTLITTEKLOOTLEEVKTLTQEKDCLKQLESQIERDQLKSDIHDVTNNMI 960 Qy

901 SPLQTVREKTLITTEKLOOTLEEVKTLTQEKDCLKQLESQIERDQLKSDIHDVTNNMI 960 Db

961 DTQEQRLNALSQKQHTINTLSKISIEVSRYNLHMEENTGETKDEFOQKQWVGIDKKQD 1020 Qy

961 DTQEQRLNALSQKQHTINTLSKISIEVSRYNLHMEENTGETKDEFOQKQWVGIDKKQD 1020 Db

1021 LEAKNTOTLTADVKNDEIIEQORKIFSLIOEKNELOQMLSVIAEKEOLKTDLENIEWT 1080 Qy

1021 LEAKNTOTLTADVKNDEIIEQORKIFSLIOEKNELOQMLSVIAEKEOLKTDLENIEWT 1080 Db

1081 TENQBELRLIGDELKQKQEIIVAQEKNHAIKKEGHSRTCDRLAEVEEKLKESQOLQEKQ 1140 Qy

1081 TENQBELRLIGDELKQKQEIIVAQEKNHAIKKEGHSRTCDRLAEVEEKLKESQOLQEKQ 1140 Db

1141 QQLLVQSEMSQKQKINEIENKVELKNKELTLEHMETERLELAQKLNENYEVKSTIK 1200 Qy

1141 QQLLVQSEMSQKQKINEIENKVELKNKELTLEHMETERLELAQKLNENYEVKSTIK 1200 Db

1201 ERKVLKEQSFETERDHLRGVIREIATGLQTKELKTAHHLKHEQHTIDELARSVSE 1260 Qy

1201 ERKVLKEQSFETERDHLRGVIREIATGLQTKELKTAHHLKHEQHTIDELARSVSE 1260 Db

1261 KTAQIINTQDLEKSHTKQEEIPVLHBEQELLPNVKVSETOETWNELELLTEQSTTKDS 1320 Qy

1261 KTAQIINTQDLEKSHTKQEEIPVLHBEQELLPNVKVSETOETWNELELLTEQSTTKDS 1320 Db

1321 TTLARIEMERLRINKEKFOESQBEIEKSLTKERNLKTITKEALEVKGHDOLKEHRETLAKIO 1380 Qy

1321 TTLARIEMERLRINKEKFOESQBEIEKSLTKERNLKTITKEALEVKGHDOLKEHRETLAKIO 1380 Db

1381 ESQSKQEOSLNKKEKNDNETTKIVSMEQPKQDSALLRIEIMGLSKELQESHDBMKSV 1440 Qy

1381 ESQSKQEOSLNKKEKNDNETTKIVSMEQPKQDSALLRIEIMGLSKELQESHDBMKSV 1440 Db

1441 AKKDDQLRLOQEVLOSSEDQKENIKIIVAKHLETEBEELVAHCCLEKEQETINELRVNL 1500 Qy

1441 AKKDDQLRLOQEVLOSSEDQKENIKIIVAKHLETEBEELVAHCCLEKEQETINELRVNL 1500 Db

1501 SEKETETSTIQOLEAINDKLQNKIOEIEYKEBQNLNKOISEVQENVNELKQFKEHRKAK 1560 Qy

1501 SEKETETSTIQOLEAINDKLQNKIOEIEYKEBQNLNKOISEVQENVNELKQFKEHRKAK 1560 Db

1561 DSALQSTIESKMLELTNRLQESQBEIQIMIKESEMKVQBALQIERDQLKENTKEIIVAKM 1620 Qy

1561 DSALQSTIESKMLELTNRLQESQBEIQIMIKESEMKVQBALQIERDQLKENTKEIIVAKM 1620 Db

1621 KESQEKYQFLKMTAVNETQEKCEIEHLKEQEPQTKNLNENIETENIRLTQILHLENLE 1680 Qy

1621 KESQEKYQFLKMTAVNETQEKCEIEHLKEQEPQTKNLNENIETENIRLTQILHLENLE 1680 Db

1681 MRSVTKERDDLSRVEETLKVERDQKENLRETTITRDLEKQEBELKIVHMLKHEQHTIDKL 1740 Qy

1681 MRSVTKERDDLSRVEETLKVERDQKENLRETTITRDLEKQEBELKIVHMLKHEQHTIDKL 1740 Db

1741 RGVISEKTNEISNMQKOLEHSNDALKAQDLKIQBELRIAHMHLKEQOETIDKLRGIVSEK 1800 Qy

1741 RGVISEKTNEISNMQKOLEHSNDALKAQDLKIQBELRIAHMHLKEQOETIDKLRGIVSEK 1800 Db

1801 TDKLSNMQKOLENSNAKLOEKIOELKANEHOLITLKKDVNETQKVSMEQOLKKQIKDOOS 1860 Qy

1801 TDKLSNMQKOLENSNAKLOEKIOELKANEHOLITLKKDVNETQKVSMEQOLKKQIKDOOS 1860 Db

1861 LTLTKSLIENLNAQELHENLEEMKSVMKERDNLRRVEETLKLERDQKESLOETKARDL 1920 Qy

1861 LTLTKSLIENLNAQELHENLEEMKSVMKERDNLRRVEETLKLERDQKESLOETKARDL 1920 Db

1921 EIQELKATARMLSKEHKEVTKLREKISEKTIQISDIQKOLDKSKDELQKIOELQKKEK 1980 Qy

1921 EIQELKATARMLSKEHKEVTKLREKISEKTIQISDIQKOLDKSKDELQKIOELQKKEK 1980 Db

1981 QLLRVKEDVNMHSHKINEMEQKKQFBNYLCKCEMONFQLTTKLHESLEIRIVAKERD 2040 Qy

1981 QLLRVKEDVNMHSHKINEMEQKKQFBNYLCKCEMONFQLTTKLHESLEIRIVAKERD 2040 Db

2041 ELRRIKESLKNRQOFTATLREMIARDQNHQVPEKELLSGQOHLMESLREKCSRIKE 2100 Qy

2041 ELRRIKESLKNRQOFTATLREMIARDQNHQVPEKELLSGQOHLMESLREKCSRIKE 2100 Db

2101 LLKRYSEMDDHYECLNLSLDLEKEIEFHRIMKKLYVLSYVTIKIKEEQHECINKFEMDF 2160 Qy

2101 LLKRYSEMDDHYECLNLSLDLEKEIEFHRIMKKLYVLSYVTIKIKEEQHECINKFEMDF 2160 Db

2161 IDEVEKQKELLIKIHOLOQDCDPSRELRDLKLNOMDLHIFEILLKDFSESEFPKIKTEF 2220 Qy

2161 IDEVEKQKELLIKIHOLOQDCDPSRELRDLKLNOMDLHIFEILLKDFSESEFPKIKTEF 2220 Db

2221 QOVLNRRKEMTOFLEEWLNTFRDIEKLNKGIQKENDRICQVNNFNNRIIAIMNESSTPFE 2280 Qy

2221 QOVLNRRKEMTOFLEEWLNTFRDIEKLNKGIQKENDRICQVNNFNNRIIAIMNESSTPFE 2280 Db

2281 ERSATISKEWQDLSKLEKNEKLPKNYQTLKTSLSAQVNPPTQDNKNPHVTSRATOL 2340 Qy

2281 ERSATISKEWQDLSKLEKNEKLPKNYQTLKTSLSAQVNPPTQDNKNPHVTSRATOL 2340 Db

2341 TTEKIRELENSLHEAKESAMHESKIIKMOKELEVNTDIIAKLOAKVHESNKCLKTKET 2400 Qy

2341 TTEKIRELENSLHEAKESAMHESKIIKMOKELEVNTDIIAKLOAKVHESNKCLKTKET 2400 Db

2401 IOVLQDKVALGAKPYKEEIEDLKMVLKIDLEKMKNAKEFEKEISATKATVYQKEVIRL 2460 Qy

2401 IOVLQDKVALGAKPYKEEIEDLKMVLKIDLEKMKNAKEFEKEISATKATVYQKEVIRL 2460 Db

2461 LRENLRSSQAOQTSVISHETDPOPSNKPLTCGGSGIIVONTKALILKSEHRLKEIKS 2520 Qy

2461 LRENLRSSQAOQTSVISHETDPOPSNKPLTCGGSGIIVONTKALILKSEHRLKEIKS 2520 Db

538	DB	KDSGMAECRKASPEKEITTSLOOQLQSKSEBEKKELVQSFELKTAELBEQLSVAKANLEMW	597
500	QY	INS-----	502
598	DB	TNSREHSINAEVQTDVEKEVVRKEMSVLGDGYNASNDIQDSSVDGKRLSSSHDECIH	657
503	QY	-----IRA	505
658	DB	RKMLEQKIVDLEEFIEINLKXSENDKOKSFEQDFMESIQLCEATMAEKANALEALMRD	717
506	QY	DYDNLVLDYEOLRTEKEKEMELKEKNDLDSEFALERKTKDOEMOLIHISINIKLVKH	565
718	DB	NFDNIIILENETLKREIADLRSKENGQTHFEILEKEKTEQHEQAQIHEIGSLKKLVN	777
566	QY	REVYNQDLENELSSKVLELREKEDIKKLOFYIDSQXLENIKMDLSYSL-----ESIEDPK	621
778	DB	AEMYNQLEEDLETCTKILKEQEIQLAELKRADNLQKKVNFNFDLSVMSGDESKLCEEIF	837
622	QY	QMKQTLDPDAETVALDAKRESAFIRSENLELEKEMKELATTYKQENDIQIYQSQLRAB--	679
838	DB	QLKQSLSDAEAVTRDAQECSFLRSNLELEKEMEDTSNWYNQKKAASLFEKQLETEKS	897
680	QY	--KMGQVDLEKELQSAFNEITKTLSDIGKVPKDLLCNLEKGIKTDLOKELMKVEENE	737
898	DB	NYKQKQADLQKELQSAFNEINYLNGLAGKVPKDLLSRVLEKKEKVSFQKLEKALEKN	957
738	QY	ALREBVILLSELKSLPSSEVERLRKRIQDKSEELHIITSEKDLFSSVHKESRVQGLLEE	797
958	DB	ALENEVTCLSYKFLPNEVECLKNQISKASEEMTLKKQEGEHSASIIISQEIIMQSQSEQ	1017
798	QY	IGKTKDOLATTONYKSTDOEFQNFKTLHMDPKQYKMWLEENRMNQELVNLKSAQKF	857
1018	DB	ILQUTDEVTHTQSKVQQTQEEQYLEMKKMHDDLPKY-----IRNKSEADLLREMNIL	1070
858	QY	DSSIGALKTKLSYKTOBLQKTRVEQERINEMQLKEQLEN--RDSPLQTVREKTLTIE	915
1071	DB	KGTWSEVVEVKIADTKHELEETIRDKQLLHEKKYFFQAMQTFPIPTPLSDSPPSKLVEG	1130
916	QY	KLQQTLE--EVKILTQEKDQLQKQSLQIERDQLKSDIHDVNNMIDTQEQLRNALES	973
1131	DB	NSDQPIEINDYHNLIALATERNNIMVCLTERNSLKEQV-----IDLNTQLQ-SLQQA	1182
974	QY	KQHQETINTLKSXISE-EVSRNLHMEENTGETKDFQKMWGDKQ--DLE-AKNTQTL	1029
1183	DB	SIEKSDLQRPQDLREGEVKLLEMLLKHGLTD---SOLSEKLOLENLEVTEKLOTL	1238
1030	QY	TADVKNELIIBOQRKIFSLIQEKNLQOQMLSEVIAFEKQKLTDLKENIEMTTIENQELRL	1089
1239	DB	QEBMKNITI-----ERNELOTFNFDLKAHDSLKQDLSENIBQSITQDELRA	1286
1090	QY	LGDELKQKEIQAQEKNAHKA-KEGELSRCTDLAEVEEKLKQKQOQKQQLINVOE	1148
1287	DB	AQELREKQQLVDSFFQQLDCSVGISPNHDAVANQEKVSGVNSQJSEM-----LRG	1341
1149	QY	EMSEMOKK----INELENLKNELKNKELTLEHMETERLELAQKLNENTVEEVKSITKERV	1204
1342	DB	ERDELQTSKALVSELELLRAHVKS-----VEGENLEITCKLNGLEKILGKSEESV	1394
1205	QY	LKELQKSFETERDLRGY-----IREIEATGLQVKEELKIAHHIKHEQHTID	1252
1395	DB	LKSMLENLKNEDNKLKEQAEEYSKKNENOFSLSEVFGSKLVDETEVLKAQLKAAEERLE	1454
1253	QY	ELRRSVSEKTAQIINTODLE-KSHTKLOKEIIPVLHEEQELLPNVKVSQETQTNNELELL	1311
1455	DB	IKDRDYFE-LVQANTNLVEGKLETPLOAD-----HEED-----SIDRRSEEM-EIKVL	1501
1312	QY	TEQSTTKDSTTLIAREMERLRNLNRPQESQEBEIKSLITKERNDLKTIKALEVKKHQLKEH	1371
1502	DB	GEK-LERNQYLLERLQEEKLELSNKLIILOKEMETSVLLKODLQKLESLSLENIILKEN	1560
1372	QY	IREFTL-----AKTQESQKQEQESLNKMKEDNTTKIVSEMQPKPKDQSAILLRIEIML	1424

5

495

495 -----

495 -----NIESE 499

Db 1561 IDTTLKHSDTAQLOKTOQLOLAKNLAIAASDNCPIOTEKE---TSADCVHPLEKIL 1617
QY 1425 GLSKRLOESDEKMSVAKEDDLDLQRLQEVLOSQSDQ-----KENIKEIVAKHLETEE 1477
Db 1618 LLTEELHOKTNEQKLEHKNELFOAQLVCEVHLMKSMISKSLSLESHKEKHDTQ 1677
QY 1478 ELKVAHCLKEQEBETINELRVNLSKE-----TEISTIQKOLEAINDKLONKIQBIYKEE 1533
Db 1678 QL-----LALAQQMVVTOEKELQQTTHLTAEDVHLKENIE-LGLNFKNEAQQKTKEQ 1732
QY 1534 QL--NIKOISEQVENV-ELKQPEHRKAKDSALQSI---ESKMLELTNRLQESQEIQI 1587
Db 1733 CLNENKELEQSQRLQCEIEELMKSLKQESALETLKESEKQVINL-----NOEMEM 1785
QY 1588 MIKEKEMKVOEALQIETDQLKENTKEIVAKMESQEB--KEYQFLKMTAVNTEQKMWCE 1645
Db 1786 VMLMEELKNSORTVIAERDQDLRESVEMSIETQDLKAKAQLQOQKDKVQELTSQ 1845
QY 1646 IEHLKEQFETQKLNLENIETINIRLTQILHENLEEMRSVTKERDDLSRVETLKVREDQL 1705
Db 1846 ISVLQEKISL-----LEN-----QWLY-NVATVKETLSERDDLQSQKQHLSEIETL 1891
QY 1706 KENURETITRDLEKQBELKIVHMLKEHOFTID---KLRGIVSEKTNIEISNMQKDLHSHN 1762
Db 1892 SLSUKEK-EFALQEAQKDK-----ADAARKTIDITEKISNTEEQLLQOATNKETL-YER 1944
QY 1763 DALKAQDLKIQBELRIAHMLKEQOFTIDKLIGIVSEKTDKLSNMQKDLSENSAKLOEKI 1822
Db 1945 ESL-----IQCKEQLALNTEHLRETLKSKDLALGKWEQERDEAANKVIALTEKMSLEEQI 2000
QY 1823 QE-----LKANEHQIITLKKOVNETQKKVSEMEQKKQIKQOQSLTISKU-----BIE 1869
Db 2001 NENVTLKEGEGEKETPYLRPSKQSSSQSSQMEELRESLTKDQLQLEAEKESISEATNEIK 2060
QY 1870 NL-----NLAQELHENLEEMSVKMERNIARVETLKLERDQKESIQETKARDLEIQ 1924
Db 2061 NLTKAISLEIEILLQNASILNEAVSERENLRHSQQLVSELEQSLTL-----KSRDHAFAP 2117
QY 1925 ELKTARMLSKHKEITVDKLRKISEKTTQISDIQKDLK---SKDELQKTKIQELO----- 1976
Db 2118 -----SKREK---DEAVNKIASLAEIKLITKEMDEFDRDSKESLQESQSHLSEELCT 2166
QY 1977 -KKELQLLR-VKEDVNM-S-HKKINMEQKKQFEPNYLCKCEMDFQTKKHLHSELEIR 2033
Db 2167 YKTELQMLKQKEDINNKLAEKVEVDEL-----LOHLSLSLEQLDQIQ 2210
QY 2034 IVAKERDELRLKESLKE-ME-RDOFIATREMIARDQNHQVKEPKRLISDGOQHLMBSLR 2092
Db 2211 MELR-NEKLRNYELCEKMDIMEKEISVLRLM-----ONEPQOE---EDDVAERMDILE 2259
QY 2093 EKCSRIKELLKRYSEM-DDHYECLNRLSLDLKEIEFHR----- 2130
Db 2260 SRNQHIELMEKISAVYSEQHTLLSSLSSELOKETEAKHCHWMLNIKESLSLTSRSPGSL 2319
QY 2131 -----IMKKLYVLSYVTKIKEBOHECINKFEMFIDEVEKQKELLIKIQL 2177
Db 2320 QTEHVKLANLTQTLANKPKVW---YRTAAVKEDHSLIKDYEKDLAAAEQKHDELRLQLQCL 2377
QY 2178 QD-----CDVPSRELRLDLKLNQMDL-----HIEILLKDFSESE---FPSIKTEFQOVL 2225
Db 2378 EQHGRKWSASAEELKFCBIEFLNELLFKKANIQSVQDDFSEVQVFLNQVSTLQEELE 2437
QY 2226 NRKEMTQPLEEWLNTFRDIEKLKNGIQKENDRICQVNNFNFRITAIMNESTEFERSAT 2285
Db 2438 HKKGFMQWLEEFGLHVDKLLSEGQWQENRIASTIQLTLKRLKAVVQSKI---QREIT 2494
QY 2286 I-SKWEQDLKSLKERNKELFRNYQTLTKSLASGQVN-----PTTODKNPHVTSRA 2337
Db 2495 VVILNPEAKLOEKKEQKELMRMRHEHGFPSASVMEENARLLGILKTQDE-----SKK 2548
QY 2338 TQLTTEKIRELENSLAEKESAMHESKLIIMQKLEVTNNDIAXI---QAKVHESNKCL 2394
Db 2549 LQ-----SRIKMLENELNDVDDAMHKEKVAILODKL-LSRNAEAEELNANQVKLTQKQDNL 2604

QY 2395 EKTERTIOVLQDKVALGAKPYKEEIEDLKMVKIDLEKMKNAKEFEKESATKATVEYO 2454
Db 2605 QAAMKEIENLQVWAGVAPYKEEIDNLKTKVKEIEMEKIKYKATQDEIAYLKSCLDK 2664
QY 2455 KEVIRLLRENLRSSQAQDTSVISEHTDPOPSNKLPTCGGSGGIVQNTKALILKSEHRL 2514
Db 2665 EGRURRKEELRRAQADNTIVCPKDYQKASFPVTCGGSGGIVQSTAMVLQSEKAA 2724
QY 2515 EKESISKQOQNEQL-----TKQNELLSNNQHL-SNEVKTWKERTLKREAHQV 2562
Db 2725 ERELSHYKKVYHLSRTMSSSEDRKKTAKSDAHSSHTGSHRSGPHKTETIR---HGPV 2781
QY 2563 TCE-----NSPK-----SPKVTGT-----ASK 2579
Db 2782 TPERSEMPSLHLSGPKKSESSTKRVSPNRSEIYSLVMSPGKTMHKLHILSPSKVLGHK 2841
QY 2580 KKOITP-----SCKERNLQDPVPKESPKSCFF 2607
Db 2842 KRALSPNRSEMPTOHVLSPGKTGLHKNLTBSTLFDNLSSPCQKQVQENL--NSPKGKLF 2899
QY 2608 DSRKSLPSHPVRVYFNDNSLGLCPFQVONAGAEVDSQPGVHASSGKDVPECKT 2662
Db 2900 DVKSKSMFY-CPSPQFDFNSKLGDFSELNTAESNDKSOAENWVWEAKKETAPCKT 2953

RESULT 3

A56539

giantin - human

N:Alternate names: macrogolgin

C:Species: Homo sapiens (man)

C>Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 10-Dec-1999

C:Accession: A56539; S37536

R:Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.

Mol. Cell. Biol. 14, 2564-2576, 1994

A:Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein

A:Reference number: A56539; MUID:94187728; PMID:7511208

A:Accession: A56539

A:Molecule type: mRNA

A:Residues: 1-3259 <SEE>

A:Cross-references: EMBL:X75304; NID:9405714; PIDN:CAAS3052.1; PID:9405715

C:Genetics:

A:Gene: GDB:GOLGB1; GCP: GCP371

A:Cross-references: GDB:454958

A:Map position: 3q13.31-3q13.31

C:Superfamily: giantin

C:Keywords: coiled coil; Golgi apparatus; transmembrane protein

F:3238-3254/Domain: transmembrane #status predicted <TMN>

Query Match 8.5%; Score 1127.5; DB 1; Length 3259;

Best local Similarity 20.9%; Pred. No. 1.2e-22;

Matches 643; Conservative 596; Mismatches 1011; Indels 831; Gaps 143;

QY 87 OTASGKTYTMMGSEDLHGVIPRAIHDFQIKKFPDREFLRVSVYMEYNETITDLLCOT 146

Db 447 ETASQTSFVDYNEGTQVTEENIASLQKRVVLENEKALLSSIE-----LEELKAEN 501

QY 147 QOKKPLIITREDVNRNVYVADLTVEVYVYSEMAKMITKGEKSRHYGEIKWQNSRSH-T 205

Db 502 EKLSQITLLEAQNRTEADREVSBISVIDIANKSSSAEES---GQDVLENTFSQKHKE 558

QY 206 IFRMILESEKCEPNSCEGSKVSHLN-----VLAGSERAQTGAAGVR 251

Db 559 LSVLLEMKBAQE-----EIAFLKQLQKRAEADHEVLDQKEMQMEGEGLAPIK 610

QY 252 LKEGCNINRSLFILGOVINKLSD-GQVGGFINYRDSKLTRILQNSLGGNPKTRIIITIP 310

Db 611 MK-----VF-----LEDTGDFPLMPNEESSLPANVEKQASTEHQSR---TSSE 651

QY 311 VSFDETLTALQFASFA-----KWKNTPTVYNEVSTDEALKRYKKEIMDLK-----KQLE 360

Db 652 ISLND--AGVELKSTNQDGKDSLSAVFDIGQCHQDE--LERUKSQILELELNFHKAQEY 707

Db 2649 SSSQKRIAELEELVVCVQKBAKKVGBIE--DKLKKELHLLHDDAGIMRNTEETAERVA 2706
Qy 2188 --LRDL-----KLNDMLHTEEI-----LKDFSEFFPSIKTFQCVLS 2225
Db 2707 ELARDLVEMQKLLMVTKENKGLTAQIOSFRSMSSILONSDHANEELDELKRYDASL- 2765
Qy 2226 NRKEMTQFLBFWLNRTRPDIKLNKGIQKENDRICQVNNFNNRIIAIMNSTEPERSAT 2285
Db 2766 --KELAQKQGL-----LNKRDALLSETAFSMN-----ST--EENSLS 2801
Qy 2286 ISKEWQDLKS-----LKEKNEKLPKNYQTLKTSILASGAQVNPFTQDNKNPHVTSRAT 2338
Db 2802 HLEKLNQQLSKQBQLHLHLSQLDESYNQVQSFKAMAS-----LONERDHLWNE-- 2851
Qy 2339 QLTTEKIRELE-----NSLHAKESAMHESKIIK-----MOKELEVT 2376
Db 2852 --LEKFRKEEGKQKRSNAQSPSTPAEVOSLKKAMSSILONDRDLLKELKNLQQYQLQIN 2908
Qy 2377 NDI--IAKQAVHESNCKLEKTKETIQVLQKVALGAKPYKEIEDLAKMLKVIDLEKM 2434
Db 2909 QETELHPLKAQLQEQY--DKTK-AFQIMQELRQENLSWQHLEHQLRMEKSSWEIHER 2964
Qy 2435 KNAKEPEKESATKATVEYQKEVIRLRENLRQQAQDTSV-ISEHTDPOPSNKLTCG 2493
Db 2965 RMKEOYLMAISDKDQQLSHLQNLRLRSS--SSQTQPLKVQYQKASPETISASP--D 3018
Qy 2494 GSGIVQNTKAL-----ILKSEHIRLEKESIKLQKQNEQLIKQKN-----ELL 2536
Db 3019 GSONLVYETELRLTQNLDSLKEIH-QKELRLIQQLNSNPSQLLEBKNTLSIQLCDTSQSLR 3077
Qy 2537 SNNQH---LSNEVTKWERTUKRAHKQVTCENSPKPKVTGTASKKKQIITPSOCKERNL 2593
Db 3078 ENQCHYGDLNHCVALKQVQELQA-GPLNIDVAFGAPQKNGVHRK----- 3123
Qy 2594 QDPVPKSPKCFPDSRSKSLPSPHPVRYFNSSILGLCPV--QNAGAESVDS----- 2644
Db 3124 SDPEELRFPQSFSEAOQLCNTQEV---NELRKLLEERDQKVAENALSVABEQIR 3179
Qy 2645 --QPGPHASSGKGVPECKTQ 2663
Db 3180 RLEHSEWDSSTPIIGSGTQ 3200
RESULT 4
152300
N: Alternate names: gcp372
C: Species: Homo sapiens (man)
C: Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 20-Aug-1999
C: Accession: 152300
R: Sohma, M.; Misumi, Y.; Fujiwara, T.; Nishioaka, M.; Ikehara, Y.
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994
A: Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized in
A: Reference number: 152300; MUID: 95100974; PMID: 7802676
A: Accession: 152300
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: mRNA
A: Residues: 1-3225 <RES>
A: Cross-references: GB:D25542; NID: g662389; PIDN: BAA05025.1; PID: g808869
C: Superfamily: giantin
Query Match 8.4%; Score 1125.5; DB 2; Length 3225;
Best local Similarity 21.3%; Pred. No. 1.3e-22;
Matches 599; Conservative 529; Mismatches 963; Indels 718; Gaps 119;
Qy 251 RLKECCNLRSLFILGVKIKSLDQGVGGFI---NVRDSKLTRI-----L 292
Db 680 KAKEISLNQ-----LIEEFKNADNNSAFTALSERDQLLSQVKELSMVTELRAQVKQL 735
Qy 293 QNSLGGNPKTRICITPVSFDETLT---ALQFASATKMYKNTPVNEV-----STD 341
Db 736 ENWLAERQRRLDYESQTAHNLNLTQIHSLSIEAKSKDVKIEVLQNELDDVQLQFSEQ 795

Qy 342 EALLKRYRKEIMDLKKOLEVSLSETRAQAMEKQDOLAELEKDLQKQVQNEKIEINTPLM 401
Db 796 STLIRISQSQONKSEVLEGAERVRHISKVEELSQALSQKEL-----EITYDQLL 848
Qy 402 VTSSS--LTIQOELKAKRKRRVTWCLGKINKMKNYSADOFNPINTITTKHLSINLLR 459
Db 849 LEKRDVETLQOITIEEKQOVTEISFMTKWQLN-BEKFSLGVEI--KTLKEQLNLLS 905
Qy 460 EIDES-----YCSSESDVFSNTLDTLSEIEMNPATKLNQENIESELNSIRADYD----- 508
Db 906 RABEAKQVEEDNEVSSGLKQNYDEM--SPAGQ-ISKEELQHEFDLLKKENEQRKRKLQ 962
Qy 509 -----NLVLDYEQLATE-KEEMELKKEKNLDLDEFALEK-----TKK 546
Db 963 AALINRKELQVRSLREELANLKDESKKBIUSETERGEVEEDK--ENKESYSEKCVTSK 1020
Qy 547 DQEMOLITHEISNLKMLVKHREVNQDLENELSSXV-----ELLREKEDQIK 592
Db 1021 QOEIY-----LQOTISEKEVELOHIRKOLEEKLAAEEQFOALVKQMNQTLQDKTNQID 1075
Qy 593 KLOEYIDS-----QKLENIKMDLSYLESIEDPKMQKQTLDEATVAL----- 635
Db 1076 LLQAEISENQAIQKLTITSNTDAS-----DGDSVALVKETVVISPPCT 1118
Qy 636 -----DAKRESAFRSNLEKELKELATTVKOMENDIOLYQSOLEAKKXQVLDLE 687
Db 1119 GSSEHWKPELEELTALKEKEQLOKLOEALTSRKAILKKAQEKERHIRELQKQDDY 1178
Qy 688 KEQSAFNEITK-----LTSIDIGKVP----- 709
Db 1179 NLEQEQDEQSKENENIGDQLRQIQVRESIDGKLFPSTDQOQSCSSPTGLEELPKATE 1238
Qy 710 -----KDLCC-----NLELEKITDLOKEL 729
Db 1239 QHHTQPVLESNCLPWFSEHSEDASALQGTSSVAQIKAKLKEIEAEKVELELKVSTTSSEL 1298
Qy 730 NKEVEENAELEEV-----ILLSELKSLPSEVERLKEITQDKSE-----ELHI 772
Db 1299 TKSEEVFQEQKQKQKQLEIESLKTWSHEAEVHAEVLSQKLESQLOIAGLEHLELOP 1358
Qy 773 ITSEKDKLFSVHVHKESRVQGLLEEIGKTKDDLTATTSNYKSTQDQFQNFKTLHMDFPQK 832
Db 1359 KLDELQKLSKKEEDVSVLSQLESE-----KEAALTQIQTIEIEQE-DLIALHTLEMQ 1412
Qy 833 YKMWLEENERNQEIIVNLSKEAQKFDSSLGALKTSELYKTOELQEKTEV-----QERLNE 888
Db 1413 AK--EHDERTIKQVQELCENKQK-PBIEIG-----BESPAKQIQKLOALISKEALKE 1464
Qy 889 MEQKLEQLENRDSPLQTVVEREKTILITEKLOQTLREEVKTLTQKDDKQKQESLQIERQK 948
Db 1465 NKSQEEELSILARG-----TIER-----LTKSLADVESQVSAQNEKKTIVLGRALLQERDKL 1517
Qy 949 KSDIHDVTNNMIDTQEOURNALLESIKQHQETINTLKSKEI SEVS--RNLHMEENTGETKD 1006
Db 1518 ITEMDSILEN-----QSLSSSCSELKALEGLTEDKEKLVEIESLSKSKIAEST----- 1568
Qy 1007 BFOQKMGVIGDKQDLKAKNTQTLADVDKDNIEEOQRKIFSLIQKNEKLQOMLESVIAEK 1066
Db 1569 EWOEKHKELOKEYELLQSYENVS-----NEAERIQHVVVEAVROEQKELYGKLRSTRANK 1623
Qy 1067 EOLKTDLKTENTMTIENOEELRLGLDELKQOEIV-AQEKNAI KKE-----GELSRTC-- 1119
Db 1624 KETEQKQLEAQOEEMEEKKWKPAK--SKQKQKILEEENDRLRAEVHPAGDTAKECME 1681
Qy 1120 -----DRLAEVEEKLKEKQOQKQOQLNVOEEMSEMKKINIEINLKNELK---NK 1170
Db 1682 TLLSSNASKMEELERVKMEYETLSKKFQSLMSEKDSLSE-----EVQDLKHQIQEIGNVSK 1735
Qy 1171 ELTLEHMETERLELAQKLNENYEEVKSIKERVKLKELQKSPETERDHLRGVIREATG 1230
Db 1736 QANLE--ATEKHD--NOTNVTGEGFQSPGTE-----BODSLMSMTR---PTC 1777
Qy 1231 LQTKKEELTIAHHLKHEQHTIDELRRSVSEKTAQIINTQDLKESKHTLOEEIPVLHEEQE 1290

Db	1778	SESVPSAKSANPAVSKDFSSHDEINNYL-----QQID-----QKERRIAGLEEKQ	1823
Qy	1291	LLPNVKVYSETQETMNELELLTEQSTTKDSTTLARIEMERLIRNEKFEQESQETSUTKE	1350
Db	1824	---KNKPSQTL-NEKNTLLSQISTKQG-ELKWLQEEVTKMLLNQOQIEELSRVTK-	1876
Qy	1351	RDNLTKIKEALEVXHDQLEKHIRTLAKIQESQSQKQSQSLNMKFKDNETTKIVSEMEQFK	1410
Db	1877	-----LKBTAEKEDKDLERELMQLAELNGISGNYQDVTDQIKNEL-----	1919
Qy	1411	PKDSALLRIETMLGLSKRLQESHDEMKSVAKEDKDLORQEVQLQESDQLEKENIKEIVA	1470
Db	1920	-----LESEMKNLKCVSLEEEKQOLVKEKTV-----ESEIRKBYLEKIQG	1962
Qy	1471	KHLE--TEELKVAHCCLEBOEBETINELRVNLSEKETEISTIQKLEAINDKLQNKIQEI	1528
Db	1963	AQEPGNKSHAKELQELLKEKQOVQKQLOKDCIRYQEKISALERTVKAL-----EF	2013
Qy	1529	YEKEEQLNKOISVEQVNVNELKQFKEHRKAKUSALQSIE-----SKMLEL	1574
Db	2014	VQTESQ---KDLKETKEN---LAQAVHRKKAQELASPKVLLDDTQSEAARVLANLKL	2067
Qy	1575	TNRLQESQERIQIMKEKEE--MKRVOEALQIERDQLEKNTKEIVAKMSESQEKYQFLK	1632
Db	2068	KKELQSNKESVKSQMKQKDEDLERLEQA--BEKHLKE-----KXMOEKLDALR	2115
Qy	1633	MTAVNETQEKMCETIEHL--KEQFETQKL--NLENIENTENIRLTQIILHENLEENRSVTKE-	1687
Db	2116	REKVH-LEETIGEITQVTLNKKDKQEVQQLQENLDSVTVQLAAFTKSMSSLDQDDRVIDEA	2174
Qy	1688	-----RDDLRSVEETLVVERDQ--LKENLRETTTRDLEKQEBELKIVMHMLKEH----	1733
Db	2175	KKWERKESDAIQSKEEBIRLKENCNSVLKQDLQMSIH-----NEELKI-NISRLHDKQI	2229
Qy	1734	-----QBTIDKLRGIVSEKTEINSMQKDLHSDNALKAQDLKIQEELRIAHM	1781
Db	2230	WESKAQTEVOLQKQVCDTLQ-----ENKELLSQLEETRHLYHSQNELAKJISELSKLD	2285
Qy	1782	HLKEQOQTIDK-----LRGIVSEKTDKLSN-----MOKDLENSNAKLOEKIOELKA	1827
Db	2286	QLTDLSNSLEKCEKQKNGLEIIRQEQEADIONSKFSVEQLETTDQASRELTSLRHEEINM	2345
Qy	1828	NEHOLLITLKQVNET-----QKKVSEMOQKKQIKDOSL-----	1861
Db	2346	KEQKILLSGKGEAIOVAIAELRQOQHDKEIKLENLLSQEENITVLEENKCAVDKTN	2405
Qy	1862	-----TISKLEIENLNAQELHENLEEMKSVMKERDNL---RRVEE---TLKLRDOLKE	1910
Db	2406	QLMETLTKIKENIQKAQALDSPKSMSSLONDRDRIVGDYQQLERHUSIILEKQDL--	2463
Qy	1911	SLOETKARDLEIOQELKTARMLSEKHEKTVDKLRKEIKETTIQISDIQKDLDSKDELQK	1970
Db	2464	-IOEAAAEENKLAEEIRGLR---SHMDLNSENAKLDAELIQYREDLNQVITIKDSQOK	2518
Qy	1971	KIOELQKKEIQLLRVKEDVNMSHKKINEMOQKKQFEPNYLCKEMDNPQLTKKLEHSL	2030
Db	2519	QLLEVQ-----LOONKELENKYLAKLEKLEKESEAN-----EDLRRSFN	2557
Qy	2031	EIRIVAKERDELRRIKESLKMERDQFIATLRMIARDRONHVKPEKRLLSPGQOHLMES	2090
Db	2558	ALQ---EEKQDLSKEIESLKVSISQL-----TRQVTAQEGVT---GLYHAOLK	2601
Qy	2091	LR-EKCSRRIKELL-----KRYSEMDDHYECLNRLSLDLEKEIEBFHRIMKKLY-----VL	2139
Db	2602	VKEEREVHRILSALPSSSQKRIAELEELVQVQEAACKVGEIE-DKLKXELKHLHDAGIM	2660
Qy	2140	SVYTKIKEEQHCINKFEMDFIDEVEKQKELLIKIHLQOQCDVPS--RELRDLKLNQM	2197
Db	2661	RNETETAER---VABLARDLVE--MEQKLLMVTKENGLTAQIQSGFRSSSL---QNS	2712
Qy	2198	DLHIEBILDQSESEPPSITKTFQOVLNRRKEMTQFLBEWLNTRFDIEKLXNGIQKENDR	2257

Db	2713	-----RDHANEELDELKRKYDASL-----KELAQKQGL-----LNRDA	2749
Qy	2258	ICQVNNFFNNRIIALIMNESFEFEERSATISKEWQDLKS-----LKEKNEKLFKNYQT	2310
Db	2750	LLESETAFSMN-----ST--EENSLSHLEKLNQQLSKDEQLLHLSQLESDSYNQVQS	2799
Qy	2311	LKTSLASGAQVNPPTTDNKNPHVTSRATQLTTEKIREL-----NSLH	2353
Db	2800	FSKAMAS-----LQNERDHLWNE-----LEKFRKSEEGKQSAAPSTSPAQVSLK	2846
Qy	2354	EAKESAMHKEKLIK-----MKELEVTNDI--IAKIQAKVHESNKKCLEKTETIQVLQD	2406
Db	2847	KAMSSQLNDRDLLEKLNQQLVYLNQBEITELHPLKAQLQEQYQ-----DKTK-AQIIMQE	2902
Qy	2407	KVALGAKPKYEELEDKMLVKIDLEKMKNAKFEKEISATKATVEYQKEVIRLLRENR	2466
Db	2903	ELRQENLSWOHELDQLRMEKSEWEIHERRMKQVYMAISDKDQQLSHLQNLIRELRSS--	2960
Qy	2467	RSQQAQDTSV-ISEHTDPQSPNKLTCGGSGIVQNTKAL-----ILKSHIRLEKIS	2519
Db	2961	SSQTPLKVQVQRQASPETSASP-----DGSQNLVYETELTQLNDSLKEIH-QKELRIQ	3015
Qy	2520	KLKQNEQLIKQKN-----ELLSNNQH---LSNEVKTWKERTLKREAHKQVTCF	2565
Db	3016	QLNSNFSQLLEKNTLSIQLCDTSQSLRENOCHYGDLLNHCVALEKQVQELQA-GPLNID	3074
Qy	2566	NGPKPKVTGTASKKKQITPSOCKERNLQDPVPKSPKSCFFDSRSKLSLSPHPVRFDN	2625
Db	3075	VAPGAPOEKNGVHRK-----SDPELRPQQSFSEAQQQLCNTRQEV---N	3117
Qy	2626	SSLGLCPYV--QVAGAESVDS-----QPGFWHASSGKQVDPCKTQ	2663
Db	3118	ELRKLEERDQVAAENALSVAEEQIRRLHESEWSSRTPIIGSCGTQ	3166
RESULT 5			
JC5837			
364K Golgi complex-associated protein - rat			
C:Species: Rattus norvegicus (Norway rat)			
C>Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000			
C/Accession: JC5837			
R/Toki, C.; Fujiwara, T.; Sohma, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.			
Cell Struct. Funct. 22, 565-577, 1997			
A/Title: Identification and characterization of rat 364-kDa Golgi-associated protein rec			
A/Reference number: JC5837; MUID:98093490; PMID:9431462			
A/Accession: JC5837			
A/Status: nucleic acid sequence not shown			
A/Molecule type: mRNA			
A/Residues: 1-3187 <OK>			
A/Cross-references: DBJ:D25543; NID:G516825; PID:BAA05026.1; PID:G516826			
C/Comment: This protein plays a role in the formation and maintenance of the characteris			
C/Superfamily: giantin			
F:49-549, 624-1176, 1238-1707, 1763-3114/Domain: coiled-coil leucine zipper #status predict			
F:3165-3187/Domain: membrane anchor #status predicted <MAD>			
Query Match 8.2%; Score 1089; DB 2; Length 3187;			
Best Local Similarity 20.9%; Pred. No. 1.2e-21;			
Matches 634; Conservative 558; Mismatches 1012; Indels 828; Gaps 124;			
Qy	113	IFQKIKKPPDREFLLRVSIMEYNETIT-----DLLCGTQMKKPLIITREDVNRNVYA	165
Db	221	VMORKLEHEEALLGRAQVDLLQKELTSAEQRNQDL--SQQLLEAEHSLRNTMEA	277
Qy	166	DLTEEVVVTSEMAKKNITKGEK-----SRHYG-----ETKMNQRSS	201
Db	278	ERQESKILMEVLEMAERKELYOLOQLERAGQAQAELEMYGTLOORHETEMEEK--	335
Qy	202	RSHTIFRMILERSKGEPSNCEGSKVSHNLVLDLAGSRAAQTGAAGVRLKSGCINRS	261
Db	336	---TACISLLQKNEQELQACD-ALKENSKLLQ-EQQQAQSAQALQOOLEDE-----	384
Qy	262	LFILQGVKKLSDGVQGVGINYRD-SKLTRILQNSLGGNPKTRIITITPVSFDELTLAL	320

385 ---LOQKSKEISO-----FVNKNLEKHEHSTSQSL-----PDVYNEGVQAV 423
 321 QPASTAKYKNTPYNEVSTDEALLARYKEINMDLAKOLEEVS-----LETRAQAMEKQ 375
 424 MEESVASLOQR-----VILENEKAGALLSSLEBELRAENEKLSRITLLEAQNAGEADG 480
 376 LA-----QLEEE---KDLQKVQNEKIENTRMLVTSSLT-----LOQEL 413
 481 MUCVESTAGTALLNRSDSSSTEESSQDVLENTFQKHKELSVLLVEMKEAQEBIAFLKSQ 540
 414 KAKRR-----RVTWCLGKINKMNS-----NVADQF 440
 541 QGRRPEGDYEVLDRKEVQMMESBGLPSVTARDVLC-----APRDKNSVPAVEGEQGRDQH 597
 441 NI-----PTNLT-----TKTHKL-----STNLTREI 461
 598 GTLEAGPLNDTGMLNSPQPDGVDKLSASPHVCLCHGELERLKTQVLELTSIHTAKET 657
 462 DESVCESDVFSNTDPTLSIEWNPATKLNQ---ENIESBLNSLRADYDNLV-----LD- 513
 658 HEKNLSEK---AKEISSITQI---TKFKESAEEARSTLTAVCEERDQLLYRVKELDV 709
 514 YEOLRTEKEMELKKEKNLDLDEPFALEKTKKQOEMLTHEISNLKNLVKHEVYNODL 573
 710 LGELRAQVRELETSIAE---AEKQRGLDYESQRAQHMLLTQIHSLSIAKSKDVKLTETL 766
 574 ENELSKVLELLREKEDQIKLOEYIDSKLENIKMDLSYLSIESIEDPKQMKQTLFDAAETV 633
 767 QRELDGVOLFSEGTQIKLSQSQIQTKE-----SEVLEGAERMK-----806
 634 ALDAKRESAFLRSENLELKEKMKELATTYKQOMENDI---OLYQSQLEAKKQOMVDLEKELQS 692
 807 -----DISKEMEELSQAQSKLEIAQMDQLLEKXKQDVET-LQOTTQE 849
 693 AFNEITKLSLIDKVPKDLLCNLELEGKITDLOKELNKEV-----EENEALREEVILLSE 748
 850 KDOQVT-----ELSPMTKVMQVLENEKFSLGVEIKTLKEQLSLLSR 891
 749 L-----KSLPEVERLRKEIQDKSEELHITSE 776
 892 AEGAKREQVEDSABSSPKPHSPHSSABEPVCKEALQOELEWLKESQRRKRLQAALIS 951
 777 KDKLFSEVYHVESRVOGLEIGTKDPLATTQSNYKSTDOEPQNFQKTLHMDFOKY---833
 952 RKELLQKVSLEBEELAKVREE---STKDSL---RESEKRELEBDSKN-----KDDPEKYGTS 1002
 834 -----KMLENERMNOEIVNLSYEAQKFDSSILGALKATELSYKTOELQEKTRV---882
 1003 EWRELEVSRLTISEKVELEGIRDLKEKAAAEELQALVQRM---TODLQNKTKQIDL 1059
 883 -QERLNEM-----EQLEQLENRDSPLQTV 906
 1060 LOEETENQATQKFIPTGMDAGDGSVAKETSVSPPRAGGHEWKPELEGK---IVDL 1116
 907 BREKTLITEKLOOTLEEVKTL---TOEKDDLKQLESQIERDQKSDIHTVNNIDTO 963
 1117 EKEKTOLQKQLEALISRKAILKKAQEKKE---KHLKEELKEQDAYR-HLQEOFDQGSKEN 1173
 964 EQLRNALSLOKHQOETINTLKSISEEV---SRNLHME-----ENTGETK 1005
 1174 ENIRAPRLQQAQESTDOOLPGTGOEPHSGEGLSLEGTEPAGESDLHAAPSPHGETA 1233
 1006 D-----EFOQKVVGID-KKQDLEAKNTQTLTADV---KUNEIIEQQKIFSLIOERNE 1054
 1234 TLQATVSVAQIQOLKEIEVEKEELEKLSIST-TSELTKSBEVLLLOEQINEQGLEIGN 1292
 1055 LOQMLESVIAEKOQKTLDKENIEMTIENTQEBRLIGDELKQOEIVAGEKHAIXKEGE 1114
 1293 LKAASHEAKAHTEQLOKEL-ESSQKIALDLEHLKTLQPELETLOKHVGQKEEVSVLVGQ 1351
 1115 LSRTCDRLAEVEKLEKESQ-----OLQEKQOQLNVQEMSEMSQKINBIENL 1163
 1352 LGEKEQTLTVQTEMEEEQBRILKAHLTQLEMQAQHEERLQVQVIEICLKQKPELE---1409

1164 KNELKNKELTLEHMETERLELAQKLNENYEEVKSITTKERKVLKELQKSF-----E 1213
 1410 -BESKAOQORLOKLAALISRKEALKENKSLQOLSDASDAVHEHLTKSLADVESQSVQN 1468
 1214 TERDHLRG-----YIREIATGLQTK-----BELKTAIHTLKEHQTIDELRR 1256
 1469 QEKDALLGKIALLOBERDKLIVEMDKSLLLENQSLGGSCSLKALGGLTEDKEKL-MKE 1526
 1257 SVSEKTAQIINTODLEKSHTKIQEETPVL-----HREOQLLPN 1294
 1527 LESVRCSTIAESTEMQEKHELOKEYEVLIIQSYENVSNEARIQHVVSVRQEKQEVYAK 1586
 1295 V-----KKVSETQETMNELE-----1309
 1587 LRSAESDKRERKQLODAEQEMEMKMKRPAKSKQOKILELEENDRDLRAEAPVYGA 1646
 1310 -----LLTBQSTTKDSTTLARIEMERLRINEXFQESQEBIKSITKERDNLTIKTALFV 1363
 1647 NESMEALLSNAVLKEE--LERITLEVKTLSKEFEALMAEKNLTSEETRNLLQVQEAQEL 1704
 1364 KH-----DOLKEHIRE-TLAKIQESQSQOESLNMKEKONETTKIIVSEMEQFPKPD 1413
 1705 QKASLETTEKSDPKDVIIEVTEAVGKSQESDLSENAKLEDAEATLANS---KP---1759
 1414 SALLRITIEMLGSKRLQESHDEMKSVAKEKDDLO-RLQEV-LOSESD---QLKENIKE 1467
 1760 -----GVSETF-SSHDDINNVLQDLQKGRIAELEMEKQKODRELSQLENEKN 1807
 1468 IVAKHELEEE-ELKVAHCCLKEQEBETINELRNLSEKETISTITQKOLEAINDKLQNKIQ 1526
 1808 ALLTOISAKOSELKL-----LEEVAKINMLNQIQEELSRTVKLTAEAEKDDLEERL-1862
 1527 EIVEKEBQNLKQISEVOENVNLEKQFKEHRKAKDSALQSTESKMLBELTNLQESQEBIQ 1586
 1863 -----MNQLAELNGSIGNYQDVTDAQIKN---EQLESEMQLKRCVSELEBKQ 1909
 1587 IMIKE-----KEEMKRVQEA---LQIERDQLKENTKEIIVAKMKESQEQYQPLK 1632
 1910 QLVKEKTKVSEIRKNEYKIQGAKPGPSKIHAKELQELKEKQEQVQKQKQKOCIRYLG 1969
 1633 MTAVNETQEKME-----TEHLKE---QFETOKLNLNENIETENIR 1669
 1970 RISALEKTVKALEFVHTESQKDLATKGNLAQAVEHHKKAQAEALSSFPKILLDDTQSEAR 2029
 1670 LTQILHNLEMRSVTHERDDLRVSEETILKVERQKLENLETITRDLQEKQELKIVMH 1729
 2030 ---VLADNL-----KLKELQSNKESIK---SQIKQK-DEDLRLRLEQAE-----KH 2070
 1730 LKEHQETIDKLRGIVSEKTNBISNMQKDLHSNDALKQAQDLKIQBELRIAHMHLKEQOBT 1789
 2071 RREKNMOEKUDALHREKA-----HVEDTL-----AEIQVSLTRKDKMKELQOS 2115
 1790 IDKLRGIVSEKTDKLSNMQKOLEN-SNAK-----LOEQLELKHANHQLITLKD 1838
 2116 LDSTLAQAAFTKSMSSIQDDDRDRVIDEAKKWEQRFDAIQTKKEBEVRLKEENCTALK-2173
 1839 VNETQKVSMEBOLKQIKQDSLTLSKLEIENLNLAQELHEN--LEEMKSVKBERDLNR 1896
 2174 -DQLRQMTIHMEELK-----ITVSRLEHD-----KEIWESKAQTELOHQOQKAYDKQE 2220
 1897 VEETILKLRDQKESQRTKARDLEIQOELKTARMLSKHEKBTVDKLR-----KI 1947
 2221 ENKELMSQLEAGQIYHSKNELTKLESELKSLKQDSTDLKNSLEKREHNNLEGIKQ 2280
 1948 SEKTTIQ-----ISDIQKOLDKSKD---ELQKIQELQKLEQLLRVKE-----DVNM 1991
 2281 QEADIQNCFNCEQLETDLTASRELTTRLHDEINVKEQKIIISLLSKEEAIQVIAIELHQ 2340
 1992 SH-KKINEMQKQKQEPFNLCBMDNFPQKTKLHESLEETIRIVAKER-----DELR 2043
 2341 QHSKEIKELNLSQEEENL-TLEENKRAVEKTNQLTALLETIKKESLEEQAKQLDSFV 2399

A;Residues: 323-387 <GUB1>
A;Cross-references: EMBL:M28266
R;Guernin-Marchand, C.; Druilhe, P.; Galey, B.; Londono, A.; Patrapotikul, J.; Beaudoin
submitted to the EMBL Data Library, April 1992
A;Description: a liver-stage-sepic antigen of plasmodium falciparum characterized by
A;Reference number: S34842
A;Accession: S34842
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 323-381, 'HKA1', <GUB2>
A;Cross-references: EMBL:M28266
A;Note: difference at carboxyl end due to frameshift error
C;Comment: This protein is found as flocculent material in the parasitophorous vacuole
C;Superfamily: trichothyalin; calmodulin repeat homology
C;Keywords: EF hand
F;154-1629/Region: 17-residue repeats (A-K-E-K-L-Q-E-Q-S-D-L-E-Q-E-R-R)

Query Match 7.9%; Score 1056; DB 2; Length 1909;
Best Local Similarity 23.4%; Pred. No. 5.4e-21;
Matches 515; Conservative 468; Mismatches 749; Indels 466; Gaps 111;

QY 492 NOENIESELSNRADYNLDVLYEQLRTEKEEMELKKEKNKDLDFEALERKTK----- 545
Db 28 NSEKDEIIKNSLRSGSSN---SNRNINEEHKKHVLHNS-----YEKTKNNENK 76
QY 546 ---KQEMQL--IHEIS--NLKVLKHREVYNQDL--ENELSSKVELLRE--KEDQIKKL 594
Db 77 FFDKDKELTMSNVKNVSNQTNFKSLLENLGVSENIFLKENKLNKEGKLIHINDDDKK- 135
QY 595 OBYIDSQKLENTIKMDLSYSLESTEDPKMQKTLFDAETVALDKRESAFRLSENLELKEK 654
Db 136 -KYIKGD-ENRQEDL-----EKAAKETL-----QGQSDLEQERL-AKEK 174
QY 655 MKELATTYKQ---MENDIQLYSQLE---AKKM---QVDLEKELQSAFNEITKLTSLI 704
Db 175 LQEQQSDSEQLAKEKLEKLEQQSDLEQLAKEKLEQQSDLEQ----- 219
QY 705 DGKVPKDLL-----CNLEEGKTLTDQKLEKNEVERNEALREEVILLSELKSLPSEVERLR 760
Db 220 ---RLAKEKLEQQSDLEQERRAKEKLEQQSDLEQERRAKEK---LQEQQSDLEQERRAK 274
QY 761 KEIQDKSELHIITSEKDKLFEVVHKESRVCGLLBEIGTKDDLTATQSNYKSTDFEQ 820
Db 275 EKLQEQQSDL-----EQERLAKEKLEQ-----OSDLEQERRAKEKLEQQSDLEQERRAKE 326
QY 821 NFKTLHMDFROK---YKMWLENERMNOBIVNLKSEAQFPDSSIGALKTSLSYKTQBLQE 877
Db 327 KLQEQQSDLEQLAKEKLEQQSDLEQ--RLAKE--KLQEQSDLEQERLAK-----E 377
QY 878 KTRVEQRLNEMQLEKLENRDPSLQTVVERKTLITEKLOQTLEEVKTLTQEKDDLKQL 937
Db 378 KLQEQQSDLEQRLAKEKLEKLEQQSDLE---QERLAKEKLEQQSDLEQERRAKEKLEQ 433
QY 938 QESLIQIER-----DQLKSDIHDVTNNMIDTQELRNALLESLSKQH---ETINTLSKIS 988
Db 434 QSDLEQERRAKEKLEKLEQQSDL-----EQERRAKEKLEQQSDLEQERRAKEKLQ 482
QY 989 EYVSRNLHME-----ENTGETKDFQKMGVIGDKKODLEAKNTQTTLTADVKDNE---IIEQ 1041
Db 483 EQQS--DLEQERLAKEKLEKLEQQSDLEQLAKEKLEQQSDSEQLERLAKEKLEQQSDLEQ 541
QY 1042 QRKIFSLIQE--KNELOQMLESVIAEK--ROLATDLD-----KENIEWTINQELRLLDGE 1093
Db 542 ERLAKEKLEQQSDLEQ--ERLAKEKLEQQSDLEQERLAKEKLEQQSDLEQERLAK 599
QY 1094 LKKQETVAQEKNAIKKEGELS--RTCDRLAEVBEKLEKESQQLQEQQLNVQEWES 1152
Db 600 LQEQSDLEQLAKEKLEKLEQQSDLEQERLA--KEKLEQQSDLEQLAKEKLEKLEQQSD 657
QY 1153 MOKKINEIENL---KNELKNKELTLEHMETERLELAQ-----KLNENYEVVKSITKERK 1203
Db 658 LERTKASKETLQEQSDLEQLAKEKLEKLEQQSDLEQERRAKEKLEQQSDLEQERRAKE 1717

1204 VLKELQSPETER---DHLRGVIREATANGLOTKBEL-----KIAHILKHEQHT 1250
718 KLQEQQSDLEQERRAKEKLQEQQSDLEQ-RAKEKLQEQQSDLEQDLRAKEKLQEQQSD 776
1251 IDELRSVSEKTAQIINTQDLEK---SHTKLQEQEIPVJHEQELLNVKVSQTQTMNE 1307
777 LEGERRA-KEKLQEQ-QQSDLEQDLRAKEKLQEQQSDLEQERRA---KEKLQEQQSDLEQ 830
1308 LELLTQSTQSTKTTLARIMERILNEXEQEQEIKSLTKERDNLKTIKEALEVHKHQ 1367
831 ERLAKEKLQEQSD---LEQER-RAKEKLQEQQSD---EQDRL-RAKEKLQEQQSD 878
1368 LKEHIRETLAKIQESQKQSLNMKEK-----DNETTIVSEMEQFKPKDSALLRIE 1421
879 LEQ-RAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSD---LEQ 932
1422 EMLGLSKRLQESHDENSVAKEDDQLRLQEVLOSE---SDQLKENIKEIVAKHLETEE 1478
933 ERRA-KEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERLAKEK- 990
1479 LKVAHCCLEQEQETINELRNLSKEFTEISTIQKLEAINDKLONKIQEIVEKEEQLNIK 1538
991 ---LOEQQSDLEQERLAKEKLQEQQSDLEQERLA-KEKLQEQQSDL---EQERLAKE 1040
1539 QISEVQENVNELOKFKHRAKQSAQSIQESKMLELTLNRLQESQEEIQIMKEK----- 1592
1041 KLQEQQSDLEQERLAKEKLQEQQSDLE-QLERLAKEKLQEQQSDLEQERLAKEKLQEQQ 1098
1593 ---BEMKRVQEAQIQLERDQKENTKEIVAKMK-ESQEKYQFLKMTAVNETQKMCETEH- 1648
1099 DLEQERLAKEKLQEQQSDLEQ-ERLAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQEQ 1154
1649 ---LKEQFETQKLNLENITENIRLTQILHENLEMRSVTKERDRLRSVEETLKVERDOLK 1706
1155 RLAKELQEQQSDLE-QLERLAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERLAKE 1209
1707 ENLRITITRDLKEQELKIVHMHKEHQETIDKLRGIVSEKTNESNMOKDLE---HSND 1763
1210 EKLQEQ-OSDLEQERAK---EKLQEQQSDLEQER-LAKEKLQEQ-OSDLEQERRAKE 1261
1764 ALKAQDLKIQELRIAHMHLKEQETIDKLRGIVSEKTKLSNMOKDLENSNAKLOEKIQ 1823
1262 KLQEQQSDLEQERR-AKEKLQEQQSDLEQER---RAKEKLQEQQSDLEQERLAKEKLQ 1315
1824 ELKANEHQILITIKDVNETQKVSEWQEKQIKDQSLTKLEITENINLAQELHENLEE 1883
1316 EQQSDLEQERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSD 1367
1884 MKSVMKERDNLRRVETLK-----LERDOL-KESLQETKARDLEIQELKTARMLSKHEK 1937
1369 QOSDLEQEQ-RAKEKLQEQQSDLEQERLAKEKLQEQ-QORDLEQERRAKEKLQEQQSD 1423
1938 ETVDKLRKISKTIQISDIQKDLQKDELQKQELQKLOLLRVKEDVNMHKKIN 1997
1424 EQERRAKEKLQEQ-OSDLEQEQ-RAKEKLQEQQSDLEQER---RAKEKLQEQQSDL- 1474
1998 EMEQLKQEPNVLCKCENDMQLTKLHESLEEIRIVAKERDLRRIKESLKMWERDQFI 2057
1475 EQERLAKE-----KLQEQQSDLEQERRAKEKLQEQQSDLEQERR- 1513
2058 ATLREMIARDNRQHVQK---PEKRLSDGQOHLMESREKSCRIKELLKRYSEMDDHYECL 2115
1514 -----AKEKLQEQQSDLEQERLANEKLQEQQSDLEQEQ-RRAKEKLQEQQSD----- 1557
2116 NRLSLDLKEHIEPHRMKKLVYLVSVTKIKEQHECINKFEMDFIDEVEKOKELIKIQ 2175
1558 -----DLEQEQ-RRAK-----EKLQEQQ-----SDLEQERRAKEKLQ 1587
2176 HLOQCDVPSRELRDLKL-NONMDLHIEITLQD---FSESEFPIKTEFOQVLSNRKEMT 2231
1588 ---EQQSDLEQERLAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERLAKEKLQEQQSD 1645
2232 QFLEEWLNTRFDIEKLNKGIOKENDR-----ICQVNN-NFFNNRIIATMNESTEF 2280

1646 DVLAEDLYGLRIPALP---LPSENERGYIIPHOSSLPQDNRGRSDSKESIIEKT---N 1700
2281 ERSATISKEWEQDLKS---LKERNKEKLFKNYQILKTSLASGAQVNPPTQDNKPHVTSRAT 2338
1701 RESIITNVGRDRDIHKGHEKKD-----GSTKPEQKEDKSDIADQNHYL 1744
2339 Q-LTTEKIRELENSLHEAKESAMHESKILIKMKQKLEVTNDIIAKLQAKVHESNCKLEKT 2397
1745 ETVNISDVNDFOISKYEDIEISAYDDSLDE-----EEDDEDLDEF 1785
2398 KETIQV-LQDKVALCAKPYKEIEDLKMVKIDLEKMKNAKEFEKEISATKATVEYQK 2455
1786 KPIQVNDQDENIGI--YK-ELEDL-----TEKNENLDLDEGI-----EK 1825
2456 EVIRLREMLRRSQOQDTSVISEHTDPQSPKPLTCGGSGIVQNTKALILKSEHIRLE 2515
1826 SSELSEKIKKGKYEKT---KNNFKENDKS-----LYDERIKKY 1864
2516 KEISKLKQNEQLIKQ-----KNELLSNNQHLNNEV 2546
1865 KNDQVNNKEKEKFIKSLFHPDGDNEILQIVDELSEDI 1902

RESULT 7
T30335
KLP2 protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 02-Jun-2000
C:Accession: T30335
R:Boletti, H.; Karsenti, E.; Vernos, I.
Cell 84, 49-59, 1996
A:Title: Xklp2, a new Xenopus centrosomal kinesin-like protein required for centrosome
A:Reference number: Z20827; MUID:96140639; PMID:8548825
A:Accession: T30335
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1388 <BOL>
A:Cross-references: EMBL:X94082; NID:g1129172; PID:e213754; PIDN:CAAG3826.1
C:Genetics:
A:Gene: klp2
C:Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology

Query Match 7.9%; Score 1048.5; DB 2; Length 1388;
Best Local Similarity 24.7%; Pred. No. 6.2e-21;
Matches 424; Conservative 271; Mismatches 561; Indels 463; Gaps 62;

QY 2 AEEGAVAVCVVRPLNSREESL---GETAQVYVKTNNVI--YQVDSKSNFDFVPHG 55
DB 22 AEEDAIVFVRIRP--PVEGTLTGVDGEGLCLTALSTTIRLHKSPEPMTFDHVAHV 79
QY 56 NETTKNVYEETAAPIIDSAIQYNGTIFAYGOTASGKTVMMSGDH-----LCVIRP 108
DB 80 DTQESVFSVAKNIVESCMNGYNGTIFAYGOTSGKTFTMLGPSNDFTNLRGVI 139
QY 109 AIHDLF---QIKKFPDREFLLRVSYMEIYNETITDLCGTQRMKPLIREDVNRVY 163
DB 140 SEYFFLLINREKEKAGEKSLKCSFEIYNEQIFDLL--DSASAGFLREHIKGVF 197
QY 164 VADLTEVYVYSEMALKWITTKGEKSHYGETKNNQSSRSRSHIFRMILRESKEGPSNCE 223
DB 198 VVGAVQVVTSAEAAYQVLSMGRNRVASTSMNRESSRSHAVFTVTTIESMEK---TNDL 254
QY 224 GSVKVSMLNLDLAGSERRAAQCAAGVRIKKEGNNRSLFILGOVTKKLSGQGV--GFI 281
DB 255 VNIRSQLNLDLAGSERQKDTQTEGVRLEKAGSINRSLSCIGQVITALVDVANGORHI 314
QY 282 NYRDSKLTILQNSLGNPKTRIICITTPVS--FDETLTALQFASAKYMNTPYNEVS 339
DB 315 CHYDSKLTFLRDSLGNKATFYIANVHPGSKCFGETSLTQPAQAKLIKNAVNE-- 372
QY 340 TDBALLKRYRKEIMDLKKLEEVSLTQPAQAMEKDQALQELLEKDLQKQVNEKIENL 399

Db 373 -----DTQGVUSQLOAEVKKL-----KEQLSQL----- 396
QY 400 MLVTSSSLTQOELKAKRRVTVCLGKINKMKNSYADOFNPTNTTTHKLSINLR 459
Db 397 -----SGOMPGDISVARVPSV-----GDNNDYWNFIAMWMLLEKSDREKKVLLQ 442
QY 460 EID--BSVCSDESDFSNLTDLSEIENWPATKLNQENISELSNLRADYDNLVDYEQ 517
Db 443 KVVQLEDLCKKKEF-----IQSNKMIKVPREDHIS----- 473
QY 518 RTEKEEMELKX- EKNDLDEFEALERKTKQDEMQLTHEISNLKULVKR- EYVNDQLEN 575
Db 474 RLEKAHKEGRISLSNNEODFIA-----ELKEERTLKQVHEHPVAKVALEN 522
QY 576 ELSKVELLEKEDQIKKLEQYIDSQ-----KLENIKMDLSYSLESTEDPKMQKTL 627
Db 523 -----HSLREENKRLHLSQSVKRAQVTAQMAEAEKAFLEVSVS- EKDRQVAPMHP 575
QY 628 FDAETVALDAKRESAFLRSENLEKEMKELATYKQEMENDIQLYQSQLEAKKMQVDLE 687
Db 576 IQLDNNSL---MSAARMREMLQLE---SELATSKOE-----YEEFKELTKKQVEQE 622
QY 688 KELQSAFNEITTKLTSIDGKVPKDLNLELEGGKITDLOKELNKEVEENALREBVILLIS 747
Db 623 SELQSLIKSNQHLNILEA-----IKANKRHEVSQALRMHAAETI 662
QY 748 ELKSLPSEVERLKEIQDSEELHIITSEKOKLFSEVHVHESRVOGLLE- EIGTKKDL 805
Db 663 KMTPTPTKSYNLSRLVPR-----LSPDAMPNGMLDTPKSGDVMDDI 704
QY 806 ATTQSNYKSTDOBFQNFKTLHMDPFOKYKQVLEENRMNOEIVNLSKEAOKFDSSGALK 865
Db 705 IN-----EPIPPEMS-----EQAYBAIAEELRIVQEV----- 732
QY 866 TELSXYTOBLOKTRVQBERLNEMLKOLENDRSDPLOTVEREKLITEKLOQTLLEVVK 925
Db 733 TALQAKLDEEGKNTLQOQVNLKELCSTQIQ-----ELFNSERSNMNKEQDILIAQIK 786
QY 926 TLTOBKDDLKQLOESQIHERDOLSKDIHDVTNMNDITQOLNANLESKQHQETINTLKS 985
Db 787 SLEKQKQENKS-----QEDVLKSEVHD-----LKVLOS----- 815
QY 986 KISEVSNLHWEENTGETKBFQOQKMGVIDKKQDLEAKNTOTLTADYK-----DNBIIRQQ 1042
Db 816 -----ADRELGAVKGEYSL-----YREKQEKLSQLSARHMDVOLQDNDVRLEHE 860
QY 1043 RKIFSLIOEKNELOQM-----LESVIA-EKEQLKTDLK-----ENIENTIENQSELRLGDEL 1094
Db 861 -----TLLEKESLQDAFNLEVMKFEIDQLKQEIISDSKHENETLRAEFSNLLELLETEK 916
QY 1095 KQOQEIQAQKHAIKKEGELSRTCDLRLAIVBEKLEKESQOLQEQOQLLVQVEMSEMQ 1154
Db 917 ERROKLTSLQEE-----DKNKTKELLOVVDENMHRLKOCSELMATKCE 959
QY 1155 KKINEIENKNEKNEKLTLEHMETERLELAQKLNENYEVKSIKERVKLKELQKSPET 1214
Db 960 QOVELHGLHSLSKSEMIADE-----KONTADVEVADLMNQIQV 1002
QY 1215 ERDLRGYRIEATGLQTKELKIAH-----LKEHQETIDELRSV 1258
Db 1003 HRTTI---IHKTESIDLLTR-----LEDIHSKYSIVLLAKESKTVIIEQEKQIBELRECL 1056
QY 1259 SEK-TAQIIN-----TOPLEKSHTKLOEIPVLH-EQOELLPNVKVSETORTM 1305
Db 1057 ERKQADNIEKELLCDDLAHATELEKLTFAFNQEAALLHTHEKEL---VEKQOISELT 1113
QY 1306 NELELLTEQSTTKDSTTLARIEMERLRINEKFPQS-----QEBIKSLTK 1349
Db 1114 NOVKLMTDLRIS-----EQEKIRPASSNSSPPVLPETPRTPGPNPVDSEIANLQK 1165
QY 1350 ERDNLTKIKEALEVKHDLKHEHRETILAKTQESQSKQEQSLNKEKKNETTKIVSEMEQF 1409
Db 1166 RNTNLEILVSELN-----EERTSKNEEIRLKMQLCETEN----- 1200

QY 1410 KPXDALLRIEIML-GLSKRLQESHDEMSVAKKDDQLQRLQVLOSEQDLKENIKEI 1468
Db 1201 -----MRLEIQNLQMGCKELKSQLENCNVMKDSNDQK-----PSDMQDLAKREIEKE 1247
QY 1469 VAXHLE-----TEBELKVAHCCLQEQETINELRVNLSEKTEISTIOKLEAINDKLQNK 1524
Db 1248 VSRMEKKGATEHILKQ-----AELEETRNIL-----CTKXHSINE 1284
QY 1525 IQBIYKBEQNLNKOISEVOENVNLKQKFEHRAKAKDSALQSIESTESKMELTNR-----L 1578
Db 1285 LSKEIERTSLRAKAFTEKEEIRSELLEGKYEEYEKLSHELDMLRKQVFLAENGKILGH 1344
QY 1579 QESQBEIQIMIK-EKEMKRVQEA--LQIERDOLKENTK 1614
Db 1345 QNPNOQIYLVKLKKNENKLEAEKLRLENLFLKESKK 1383

RESULT 8
T08621
centroome associated protein CEP250 - human
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T08621
R:MacK, G.J.; Rees, J.; Sandblom, O.; Balczon, R.; Fritzler, M.J.; Rattner, J.B.
Arthritis Rheum. 41, 551-558, 1998
A:Title: Autoantibodies to a group of centrosomal proteins in human autoimmune sera react with CEP250
A:Reference number: Z16462; MUID:98165428; PMID:9506584
A:Accession: T08621
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2442 <MAC>
A:Cross-references: EMBL:AF022655; NID:g2832236; PIDN:AAC06349.1; PID:g2832237
A:Experimental source: cell line HeLa

Query Match 7.7%; Score 1026; DB 2; Length 2442;
Best Local Similarity 20.5%; Fred. No. 4.2e-20;
Matches 555; Conservative 561; Mismatches 945; Indels 642; Gaps 109;
279 GFINYRDSKLRILQNSLGGNPKTRICTITPVSPDELTALQFATAKYMKNTYVNE- 337
7 GLNNMKPQSLQVLVEEQV-----LALQQMAENQAASWRK-LKNSQEAQOR 51
338 -----VSTDEALLKRYKREIMDKKQL-----EVSLETQRAQAMEKQD----- 375
52 QATLVKRLQAKVQVRSWCQLEKRLKLEATGGPIQORWENVEPNLDELVLREEQORCE 111
376 -LAOL-----LEEKDLQKQVNEKIENTL-----RMLVTSSSLTQOE-----LK 414
112 SLAEVNTQIRLHMEKADVVKALRADVEKLTVDWSRARDLMRKESQWMEQEFPKGYLK 171
415 AKRRRVTVCLGKIN-----KMKNSVADQFNPTNTTTHKTH-KLSINLRLEIDESVCS 467
172 GEHGLLSLREVVTFRRHFLEKSAITDRDLMELKAE-----HVRLSGSL-----TCCL 221
468 ESDVFSNLTDLSEIENWPATKLNQENISELSNLRADYDNLVDYEQRLTEKEEMELK 527
222 RLTVGAQR-----EPNGSRMDGRE--PAQLLLLLAKTQEL-----EKEAHERSQELIQ 269
528 LKEKNLDLDEFEALERKTKQDEMQLTHEISNLKULVKRREYVNDQLENLSKVELLEK 587
270 LKSGDLEKALQDR-----VTLSALLTQSQKONEDYEKMIKA----- 308
588 EDQIKKLEQYIDSKLENIRKMDLSYSLESTEDPKMQKTLFDARTVAL---DAKRESAFL 644
309 ---LRETVEILETNHTELMHEHSELSNAQOEKLSQOQVTKDITQVMVERGDNTAQSG 365
645 RSNLELEKEMKELATYKQEMENDIQLYQSQLEAKKMQVDLEKELQSAFNEITKLTSLI 704
366 EN-SLEESSIFS-QPDYQADAKALTVRSVLTRRQAVQDLQQLAGCQAEVNLQOQH 423
705 D-----GKVPKDLCCNL-----ELEKIDTLOKE---LNKEVEENALREEVILLSELKS 751

Db 424 DOWEEGKALRQLOKLTGERDTLAGQTVLDQGEVDSLSKERELLQKABEE--LRQOLEV 481

Qy 752 LPSEVERLRK---EQ---DKSELIHITSEKDKLSEVVHKEGRVQCLLEE 797

Db 482 LEQEAWLRRVNVVELOQSDSAQOKEQOEELHIAVRERERLOEMLGLEAKQESLSE 541

Qy 798 IGTGKDDIATQSNYSKTQEQFQNFKTLHMDPEQYKMYLEENRMOBIVNLSKEAQKF 857

Db 542 LITLRALESIIHLEGELLQEQTEVTAALARAQSOIAELSSSENTLKTADVADLRAAAVKL 601

Qy 858 DSGSLGAL---KTELSYKTQELQKTRVQERLNMELQKEQLERNRSPQTVREKTLIT 914

Db 602 SALNEALALDKVGLNQQLLEBENQSVCSMRBAEQARNALQ---VDLAEBKREALW 658

Qy 915 EK---LOOTLEEVKTLTOE-KDILKQLOESLOIERDQLKSDIHDVTNMNIDTQELRNAL 970

Db 659 EKXTHLEAQLOKABEAGAELOADRIDQER-----KEEIOKLSESHPQEAATQ 710

Qy 971 ESLKQHOETINTLKSITSEBV-SRNLHMEENTGETKDEFOQKMGVGD-KODLEAKNTQT 1028

Db 711 EQL---HQE-----AKRQEEVLARAVQKEALVREKAALVRLQAVERDRQDLAAQ-LOG 761

Qy 1029 LTA-DVKDNEIIEQQRKIFSLIQKNELOOMLESVIAKEQLKTDLKENIEMTIOQEE 1086

Db 762 LSSAKELLESSLFEAQOQNSVIDEPOQOLEVOIQTVTQAEVI-----QGE 807

Qy 1087 LRLIGDELKQOEIVAOEKHAIKKEGELSRTCDRLAEVVEKLEKESQOQLEKQOQLNV 1146

Db 808 VRCLKELDTESQAEQERDAAAR-----QLAQAEQGTALQOQKAAHEKENVQ 857

Qy 1147 QEEMSEMOKKINETENLKNELKNKELTLEHMETERLELAQKLNENYBEVKSITKERVILK 1206

Db 858 LREKWEKERSHQELAK-----ALESLEKWELEMLKBEQOTEMEALQAREER 909

Qy 1207 E-----LOKSFETERHLRGYIREIATGLQTKBELKIAHIHKEHQETIDELRRSVSE 1260

Db 910 TOAESALCOMQLETEKERV-----SLLETLLQTKELADA-----SQOLERL 952

Qy 1261 KTAQIINTOLEKSHTKLOEIPVLHBEQELLPNVKVSETQETWNELELLTEQ----- 1314

Db 953 -----QDMKVQKLEHQTGILQ-----TQLOEAQRELKEAARQHRDLA 992

Qy 1315 STTKDSTTLARIEMERLNEKFORSEIEKSLTKERNLNTIKIEALEVKHDQLEHIRE 1374

Db 993 ALQOESSLLQKMD-----LQKQVEDLKSQVLAQDDSQLVE-----QEVOEKLE 1039

Qy 1375 T-LAKIQESQKQESQSLMKEKNETTKIV-SEMEQFKPKDSALLRIEIMELGSLKRLQ 1431

Db 1040 TOEYNRIQKLELEREKASITLSIMEKEBQRLVLQOEAADSIRQOELSALR-----QDMQ 1090

Qy 1432 ESHDEMKVAKEDODLQRLQEVLOSQDLEN-----IKBIVAKHLETEBEELKVAHCCLKE 1488

Db 1091 EAQGEKLSAQMEILR---QEVKEKADFLAQEAQLLEELASHI-TEQOLRAS---LWA 1144

Qy 1489 QEETINELRVNLSEKETEISTI-----QKLEAINDKLQNKIETIYKEBQI- 1535

Db 1145 QEAQAQLHLRLRSTESQLEALAAQQPCGNOQAQAQALASLYSALQOALGSCYCEPPELS 1204

Qy 1536 -----NIKQISEVOENVNELKQKEHRAKDSALQSLKEMLETLNRELQSEBQIOM 1588

Db 1205 GGGDSAPSVGLEPQNGARSL---FKRGPLLALTALSAABASALLHQDLQWTKQTRDVL 1262

Qy 1589 IKEKEMKVOEAL---QIERDQKENTKEIVAKMKESQEKYQFLKMTAVNETQEMKCE 1645

Db 1263 -----RDQVQKLEERLDTTEAKSQVHTELQDLQRLQSLQONQBEKSKW-----EGKQNSLE 1312

Qy 1646 IEHLKEQFET-----QKLENLEIETENIR-LTQILHLENL-----EEMRSVTKE 1687

Db 1313 SE-LMELHETMASLQSLRLARAELOEQAQERELLQAAKENLTAQVEHLQAAVVEARAQA 1371

Qy 1688 -----RDLRSVEETLKVREDQKLENRETTITRDLQEKBELKIVH-----MH 1729

Db 1372 SAAGILEEDLRTARSALKLNKEEV-ESERER-AQALQEQGELKVAQKALQENLALLTQT 1429

Qy 1730 LXEHOETIDKLRGIVSEKTNESNMOK-----DLEHSDALKQAQDLKIOE----- 1774

Db 1430 LAAREEEVETLRGQIOLEKQ-REMQKALELLSLLDKKQENQVQDLQOQIOLEKCRSV 1488

Qy 1775 -----ELRIAHMLKEQOETIDKLRGIVS-----EKTDKLSNMOK-----DL 1811

Db 1489 LEHLPMVAVQERQKLTQVREQIREPEKDRQETQVLEHQLLELEKQDMIESQORGQVQDL 1548

Qy 1812 ENSNAKLQKIOELKANEHQLITLKKOVNETQ-----KKVS-----EMEOLKQIOKQDS 1860

Db 1549 KXQVLTLECALALEENHMKECQOKLKELEQORETORVALTHLTHLIDLEERSQEIQAQS 1608

Qy 1861 LITSLKLEIENLNAQELHENLEMKSVKBERDNLRVEE---TLKLERDQKESLOFTKAR 1918

Db 1609 SQIHDLESHTVLARELQERDQEVKSQREQIELOKQKEHLTQDLERRDQELMLQKERIQ 1668

Qy 1919 DLEIQQELATARMLSKEHETVDKLRKISEKTIQISDIQKDLKSDKDELOKQIOELQK 1978

Db 1669 VLEDQRTQOT-KILEEDLEQIKLSLRGREGRELTQRLMQERABEGKPSKAQRGSLH 1727

Qy 1979 ELQLLRVKEDVNMHSHKINEMEQKQFEPNYLCKCEMDNFQTLTKLHESLEIRIV--- 2035

Db 1728 KLILRDKKEVECCQEHILHQBELKQOLE-----QOGLHRKVGETSILLSQ 1775

Qy 2036 -----AKERDELRRIKESLSKMERQFIATLEMIARDRQNHQVPEKLLS 2081

Db 1776 REOEIVVLQOQLEAREQOGLK---EQSLOQLDE---AQALAQRDQLEALQOEOQ-QA 1829

Qy 2082 DGOQHLMESLRKCSRIK-----ELLKRYSEMDDHYECLNRLSLDLKEKEFHRIMK 2133

Db 1830 QOQE---ERVKEKADALQGALEQAHTLTKERHGBELQDHKEQARLEBELAVE---GRVQ 1883

Qy 2134 KLYVLSYVTIKKEBQHECINKPEMDFIDEVQK-----KELLIKIHLQOQCDVPSR 2186

Db 1884 ALBEVLGDLRAESREQEKALLAQOQCAEQAEHEVETRALQDQSWLQAOVLKED---Q 1940

Qy 2187 ELRDLKLNQNDLHIEILKDFSESEPPSITKTEFQVLSNRKEMTQFLBEWLNTFRDIEK 2246

Db 1941 ELEALRAESSSRHQEAAARAEAE-----LQEALGKAHAALQKEQHILLEQAELSR 1992

Qy 2247 LKNG-----IQKENDRIQVNNFNNRIIANNESFEERSATISKWEODLSIKENNE 2302

Db 1993 SLEASTATLOASLD-ACQASHRQLEEARLI---QEGEIQDQDLRYQEDVQOQQAQAQDE 2049

Qy 2303 KL---FKNYQTLKTSLSGAQVNETQDNKNPHVTSRATQTLTKERLENSLHAKESA 2359

Db 2050 ELRHQEREQLLEKSLAQRVQEN-MIQEKQN-----LGLEREEIRGLHQSVRLQTL 2103

Qy 2360 MHKESKIIKMOKELEVTNDIITAKLOAKVHESNKC---LEKTKETIQVLDQKVALGAKPYKE 2417

Db 2104 AQKQOEIILDELTQOQNN---LEALPH-SHKTSPEEQSLKLDLSLEPLR-----Q 2150

Qy 2418 BIEDLKMVLVI---DLEKMKNAKEFEKEISATKATVYQKEVIRLLR-----ENLRQOQ 2470

Db 2151 ELERLOAALRQTEAREIEWREKAQDLALSATQKASVSSLOEVAMFLQASVLERSEQOR 2210

Qy 2471 AQDTSVISE-----HTDQPSNKPITCGGSGIV-----QNTKALIL 2507

Db 2211 LODELELTRALKEKRLHSPGATSAELGSRGEQVQIGEVSGVABSPDGMKQSWRQ 2270

Qy 2508 KSEHI-----RLEKISKLKQONEQOLIKQKELLNNQHLNNEVKTWKERTILKREAHK--- 2560

Db 2271 RUEHLQOAVARLEIDERSLORHNVO-----LRSTLEQVERE-----RRKLKREAMRAA 2318

Qy 2561 -----QVTCENSPPKPYTGTA 2577

Db 2319 QAGSLEISKATASSPTQODGRGKNSNAKVAELOKEVVLLQAOQLTLERKQODYITRSA 2378

Qy 2578 SKKKQI-----TPSQCKERNLQDPVPKESPKSCFSDRSKSLPSP 2617

Db 2379 QTSRELAGLHLSHLSLAVAPAEATVLEAETRLD-----ESLQTSITSP 2425

Query Match	7.4%;	Score	980.5;	DB	2;	Length	1780;
Best Local Similarity	22.3%;	Pred. No.	4.8e-19;				
Matches	425;	Conservative	367;	Mismatches	605;	Indels	511;
						Gaps	74;
QY	9	VCVVRPLNSREESIGETAQVTKTDDNNVIYQV	-----DGSKSFNEDRVFH	54			
Db		: :	: :				
		: :	: :				
		: :	: :				
61	VCLIRPPTQSEKESECGVHILDSQTVLKEPQICILGRLSEKSSGMAQKFSKVFG	120					
QY	55	GNETTKVNYEEAALPILDSATQGYNGTTFAYQQTASGKTYTMMGSEDLGVTPRAIHIF	114				
Db		: :	: :				
		: :	: :				
		: :	: :				
121	PATTOKEPFGGICMOPVKDLKGQSLFTYGLTNSGKTYTFOGTEENIGILPTLNVL	180					
QY	115	-----QKIKKFPDREF	-----LLR	-----128			
Db		: :	: :				
		: :	: :				
181	DSLQRLYTKMLKPHRSREYLRUSSEQKEIASKALLRQIKVTVHNSDDTLYGSL	240					
QY	129	-----VSYMEIYNETITLLCGT	-----QKMKPL	152			
Db		: :	: :				
		: :	: :				
241	TNSLMSIEFEESIKDYEQANLNMANSIKFSVWVFFEIYNEYIYDLFPVFSKFKQKML	300					
QY	153	LIREDNERNVTADLTTEVVVTSEWALKWITKGKSRHYGTGMNQRSRSHITFRMILE	212				

Db	301	RISQVGVGKFKDQWQVSDSKAYRLKLGIRHQSVAFKLNASSRSHSIPVTKIL	360	1244	DYADLKEKLTDAKKQIKQVQKEVSVNRDEDKLLRIKINELEKKKQCSQELDMKORTIOQ	1303
Qy	213	SREKGEPSNCEGSKVSHNLVDLAGSERAAQTGAAGVRLKGGCINIRSLFILGQVKKL	272	1177	METERLELAQKNE---NYFEV-KSITKERVVLKEL-----QKSFETERDH-LRGYIRE	1225
Db	361	QIEDSEMSRV---IRVSELSCLDLAGSRTMTQNEGERLRTGNTINTSLTGLKCNVL	417	1304	LK-EQLN-NQKVEEAIQYERACKDLNVKEKTIEDMRMTLEBQEQTVQDDQVLEAKLEB	1361
Qy	273	SGQVGGF---INRDSKLTIRLONSLGNPKTRITCTTPVS-----FDELTALQFAS	324	1226	IB--ATGLQT-KE-----ELKTAHILKHEQHTIDELRRSVSEKTAQIINTOD-LEKSH	1276
Db	418	KNSEKSKFOQHVPFRSKLTHYFQSPNGKGK---ICMIVNISQCVLAYDEILNVLFKSA	474	1362	VERLATELEKWEKKNDELTKNQRNSKHEHNTDVL-----GKLTNLQDELQEQE	1413
Qy	325	TAKYMKNTPVNEVSTDEALLKRYRKEIMDLKQLEVESLET-----RAQAMEKD	374	1277	K-----LOBEIPVLHE--EQELLPN--VKVSVETQ-----TMELELLTQOSTTKDS	1320
Db	475	IAQ-----KVCVPDT-LNSSQDKLFGPVKSSQDVSLSDSNSKILNVKATISWEN	524	1414	KYNADKKWLEKMWLITQAKAENIRNKEMKYAEDRERFFKQOQEMELTIAQLTEKDS	1473
Qy	375	QLAQLLEKDLQKQVN-EKIENL-TRMLVTSSLTLOQB---LXAKRRERVTCWLGKIN	429	1321	TTLARIEMRLRLNEKFOESQEBEISLTYKERNLKITKEALEVKHDLQKEHIRETLAKIQ	1380
Db	525	SLEDMEDEDLVELENAETQNVETKLLADEDLDKLEENKAFISHEEKRAKLDLIEDLK	584	1474	-----DLQWREERDQVLA---ALEI---OLK-----ALLS	1498
Qy	430	KM---KNSYADQFNIPNTITTKTKLSINLAREIDESVCSSEDSVFSNTDLTSLSEIWN	485	1381	ESQKQEQSLNKKEDNNTKIVSEMEQKPKDSALL---RIRIEMGLSKRLQESHDEM	1437
Db	585	KCLINEKKEKLTUEFKLIREEVTQFTQYWAQRAADFKETLLQREILEENAEARLAFKD	644	1499	SNVOKDNEIQLKRIITSETSKITETQIMDKPKKISSADPKLQTEPLSTSEISRNKIED	1558
Qy	486	PATKLLNQNIESIELNSLRADYN---LVLDTYEQLTE-----DOEQMLIHEISNLKN	561	1438	KS-----VAKEKDQLOR-----LQEVLOSES	1458
Db	645	LVGKCDTREAAKDI CATKVTEBEATACLEKFNQKAEKAKTKGELIKTKBELKKRENE	704	1559	GSVVLDSCEVSTENDQSTRFPKPELEIQFTPLQPNKMAVHPGCTTPTVTVVEIPKARKRS	1618
Qy	533	DLDFEALERTKK-----DOEQMLIHEISNLKN-----	561	1459	DQKKE-----NIKEIVA---KHLETEBELKVAHCCLEKEQETINELR---	1497
Db	705	SDSLIQLETSNKKIITQONRIKELINIIDQKEDTINEFONLKSHMENTFKCNDAKATSS	764	1619	NEMEEDLVKCNKKNATPRTNLKFPISDDRRNSSVKKEQKVA---IRSSKKTYSLRQAS	1675
Qy	562	LVKHRVYNQDLNENLSSVEL-----LREKEDQIKKLQBYI-----DSQKL	603	1498	---VNLSEKETEISTTQK---OLEAINDKLQNKIQEIEYKEEQLNKOISEVOENNELK	1551
Db	765	LIINNKLICNETVEVPKDSKICSEKRVNENELQODEPPAKKGSIHVSSAITDQKS	824	1676	IIGVNLATKKKE-GTLQKFGDFLOHSPSILQSKAKKIETMSSSKLSNVEASKENVSPK	1734
Qy	604	ENIKMDLSILESIEDPKQKQTLFDATVALDARESFAFRLSENLEKKEKELATYK	663	1552	QFKEHRKAKDSALQS---IESKMLELNLNLOESQEEIQTIMKEEENK	1596
Db	825	EEVFPNIA---ETEDIRVLOE-----NNEGLRAFLTTIENELKNEKEEKAELNK	870	1735	RAK---RKLYTSEISSPIDISQVILMDQMKKESDHQI---IKRLRTK	1777
Qy	664	QMENDIOLYQSLBAKKMQQVLDLEKELQSAFNEITKLTSLIDGKVPKDLLNLEGGKIT	723	RESULT 11		
Db	871	QIVH---FOELSLSEKKNLTLSKEVO-----QIQSNYDITAEAL	907	D96619		
Qy	724	DLOKELKEVENEALREEVILLSELKSLPSEVERLKEIQDKSEELHIITSEKDKLFS-	782	C:Species: Arabidopsis thaliana (mouse-ear cress)		
Db	908	HVQSKNQEQEE-----KIMLSNEIETATRSITNNVSQIKLMTKIDELRTL	955	C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001		
Qy	783	EVVHKESSVOGL-LEEI--GRTKDDLATQSNYSKTDQEFONFKTLHMDFPQKYKMWLEE	839	C:Accession: D96619		
Db	956	DSVQSISNIDLLNRLDLSNGSEEDNLPNTQLDLLGND-----YLVSQVKEYRIQE	1006	R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso		
Qy	840	NERNMQEIVNLSEKAKQFDDSLGALKTELSTYQLOEKTREVOERLNMEMOLKOLENR	899	Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.		
Db	1007	PNREN-----SPHSSIEAT-----WEECKEIVKASSKSHQOIELEQOIEKL	1048	ansen, N.F.; Hughes, B.; Huizar, L.		
Qy	900	DSPLQTVREKTLITELKQTLLEEVKTLTQEKDDL-KOLESLOIERDQLKSDIHDVTVM	958	Nature 408, 816-820, 2000		
Db	1049	QAEVKGVDENNRKKEHKQDD---LLKKEETLIQQLKEELQ-----EK	1091	A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.		
Qy	959	NIDTQOELRNAME---SLKQHQETINTLKSGTSEVSRNLHMEENTGTYKDFQKMWGI	1015	C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali		
Db	1092	NVTLDVQIHVVEGRALSELTOGVTCYKAIKE-----LETI	1129	Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.		
Qy	1016	DKKQDLKAKNTQTLTADVKDNEIIEQORKIFSLIQEKNELQOMLESVIAEKQL---	1072	A:Authors: Salzbeg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,		
Db	1130	LETQKVERSHSAK-----EQDILEKSEITILKERNLKEFQHLQDSVKNQKDLNVKELK	1184	ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.		
Qy	1073	LKENIEMTTEOEBELRLGDELKQOEIVAOBKHAIKKE-----GEISRTCD	1120	A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.		
Db	1185	LKEETITLTNNLQDMKHL-LQLKEEBEETNRQETKLEELSSASSARTQNLKADLQKKE	1243	A:Reference number: A86141; MUID:21016719; PMID:11130712		
Qy	1121	RLAEEVEELKESQLOEQKQQLNVQOEMSEMOKKINEIENKN-----ELKNKELTLEH	1176	A:Accession: D96619		
				A:Status: preliminary		
				A:Molecule type: DNA		
				A:Residues: 1-888 <STO>		
				A:Cross-references: GB:AE005173; NID:g8778739; PIDN:AAF79747.1; GSPDB:GN00141		
				C:Genetics:		
				A:Gene: T30E16.9		
				A:Map position: 1		
				Query Match 7.3%; Score 971.5; DB 2; Length 888;		
				Best Local Similarity 31.2%; Pred. No. 4.2e-19;		
				Matches 301; Conservative 149; Mismatches 305; Indels 209; Gaps 30;		
				Qy 7 VAVCVVRPLNRSRLSGETAQVYWKTDNNVI-----Y 39		
				Db 4 ICVAVRVRP-----PAPENGASLWKVEDNRISLHKSILDTPTTASHAFVSGISISTDLI 57		

QY	40	QVDGSKSNP-----DRVPHGNETTKNYIERAIIPIIDSAIQVNGTIFAYGOTA	89
Db	:	:::	:
Db	58	EIVSLFLFSGVVGYFFLLPADHVFDESSINASVYELLTKDIIHAAVEGFNGTAPAYGOTS	117
QY	90	SGKTYTMGSSEDLGVIPRAIHDIPOKIYKFPDPREPLLRVSMEIYNITITDLLCGTQM	149
Db	:	:::	:
Db	118	SGKTFTMTGSETDGPLIRKSVRDVFERIMI SDREFLIRVSMEIYNEBINOLL--AVEN	175
QY	150	KPLIIREDNRRNVVADLTVEEVVYSMALWITKCEKSRHYCTEKNQRSRSHITFRM	209
Db	:	:::	:
Db	176	ORLOIHEHLERGVPFVAGLKERIVSDAQILKLIDGSEVNRHPGETNMNVHSSRSHTIFRM	235
QY	210	-----ILBSREKGPSNCEGSYKVXSHLNLVDLAGSERAQTGAAGVRLKEGCN	257
Db	:	:::	:
Db	236	VRFRSYERDLLLVIESR--GKDNSSDAIRVSVLNVLVLAGSERIAKTGAGVRLQBGKY	293
QY	258	INRSFIILGOVIKKLSDG-QVGCFPINYRDSKUTRIIONSLGNPKTRIICTTPVS--PD	314
Db	:	:::	:
Db	294	INKSMILGNVINUKUSDSTKURAHITPYRDSKULTRILOPALGCNAKTICTIAPEEHHIE	353
QY	315	ETLTALOPASTAKYMKNTPVYNEVSTDEALLKR YKETMDLKKOLEEVSLETRAQ----	369
Db	:	:::	:
Db	354	ESKGTITPASRAKITNCAQVNEILTDAALLKRQKLEETEUMKLOGSHAEVLEOEILNL	413
QY	370	-----AMEKDQAQLLEEKDLLQKVONE-----KIENLTRMLVTSSSILTLQOELKA	415
Db	:	:::	:
Db	414	SNOMLKYELECERLKTQLEEEKRKQKECENCIKEQOMKIENLNN-FVTVNSDF-----	464
QY	416	KRKRVTWCLGKINMNKNSYADQNIPNTNITTTHKLSINLLRIDESVCSESDFVSNT	475
Db	:	:::	:
Db	465	KRNQSEDITIISRKPTDGLCNVNDTSDEVGTCFKFSASRSPVVARSNYSGLSDFPSMVHS	524
QY	476	L-----DT-----LSETEWMPATKL-----LNQENIESELMSLRADY	507
Db	:	:::	:
Db	525	LGDVADEDTWMLNKGFVADDQIQTFPAVKCQPPTSIATECPREN-HSEVEDLKS--	581
QY	508	DNLVLDVEQURTEKEEMELKLEK-----NDLDEFALERKTKKQDEM-OLIHIEINLK	560
Db	:	:::	:
Db	582	-----RIQLTTNENDSLQVKFNEQVLLSNLMOEMS ELKQETLTVKEIPWRKSESVANCK	636
QY	561	NLVKHREVYNQDLENELSS-----KVBLLEKEDQIKKLOYIDSQKLENIKM	608
Db	:	:::	:
Db	637	DVYKDVITVMKSLITDKESPANTALLGTTTITTSLLATLETQFSMI--MDGQKTGS---	690
QY	609	DLSYLSLSIEDP-----KOMQTLFDFAETVAL-DAKESAFLSENLELEKMKELATTY	662
Db	:	:::	:
Db	691	-----SIDHPLSDHWETUVRVNJNTTTTTLLSDAQAKDEFILSHN-----	729
QY	663	KQMENTIQLYQSOLEAKKMQVLEKELQSAFNIEITK---LTSIDGKVPKDLLCNLELE	719
Db	:	:::	:
Db	730	-----KLDVVVRLLFHL--SLRQVLNVILKQETALEEKKLASELIU----IK	772
QY	720	GKITDLQKELANKVEENEALRE-EVILLSELKSLPSVEVERLRKEIQDKSELHIITSEKD	778
Db	:	:::	:
Db	773	ERYNELEKELCLQKOLLEASRESHEKLIKVEQFLKBERDSLDRKISQSTQRJARVIASDXE	832
QY	779	KLFSEVHKRESRYOGLEEIGTKDDLATQSNVYKTDQBFQNFKTLHMDFEOKYMWLE	838
Db	:	:::	:
Db	833	NALKDLNVKRRKDMEEETIKHISIAPAT-----RHKS FVSFHSEIKSMQKLAIT	882
QY	839	ENER 842	
Db	:	:::	:
Db	883	QNSK 886	

RESULT 12
A54803

microtubule-associated motor KIF4 - mouse
N/Alternate names: kinesin-related protein KIF4
C/Species: Mus musculus (house mouse)
C/Date: 05-Apr-1995 #sequence_revision 05-Apr-1995
C/Accession: A54803; D44259
R/Sekine, Y.; Okada, Y.; Noda, Y.; Kondo, S. A.

J. Cell Biol. 127, 187-201, 1994
A>Title: A novel microtubule-based motor protein (XIF4) for organelle transports, whose
A/Reference number: A54803; MUID:95014709; PMID:7929562
A/Accession: A54803
A/Molecule type: mRNA
A/Residues: 1-1231 <SEK>
A/Cross-references: GB:D12646; NID:G563772; PIDN:BA02167.1; PID:d1002657; PID:G563773
R/Aizawa, H.; Sekine, Y.; Takemura, R.; Zhang, Z.; Nangaku, M.; Hirokawa, N.
J. Cell Biol. 119, 1287-1296, 1992
A>Title: Kinesin family in murine central nervous system.
A/Reference number: A44259; MUID:9307686; PMID:1447303
A/Accession: D44259
A/Status: nucleic acid sequence not shown; not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 91-111, 'S', 113-240 <AIZ>
A/Experimental source: brain
A/Note: sequence extracted from NCBI backbone (NCBIP:118904)
C/Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology
K/Keywords: ATP; nucleotide binding; P-loop
F/18-343/Domain: kinesin motor domain homology <RMOT>
F/88-95/Region: nucleotide-binding motif A (P-loop)

Query Match 7.1%; Score 952.5; DB 2; Length 1231;
Best Local Similarity 26.0%; Pred. No. 1.8e-18;
Matches 367; Conservative 223; Mismatches 491; Indels 329; Gaps 51

Qy	7	VAVCVVRPLNGREESLG-ETAAQVYWKTDNNVIYOVDGSKSFNDFRVPFHGNETTKNVYEE	65
Db	10	VRVALRCRPLVSKIEKGQCOTCLSPVGPQVV--VGNDKSYTFDFVPDSTQBEVFWT	67
Qy	66	IAAPIIDSAIQYNGTIFAYGOTAGSKTYMMGS-----EDH---LGVIPRAIHIDFQIK	118
Db	68	AVAPILKGVFGKYNATVLAIGYGTGSKTYSMGCAVTAEQHDSALGVI PRVIQLLPKEIN	127
Qy	119	KFPDREFLLRVSIMEIYNETTDLCCGT-QQMKPLI IREDVNRNVYVADLTBEVYVTSM	177
Db	128	KSDREFLLKVSYLEIYNEEILDLCSSREKATQINIREDPKEGKIVGLTEKTKTVLVSAD	187
Qy	178	ALKWTTKGEKSRHYGETKMNORSRSHITFRMILESRKSGPCNCEGSKVSHLNLVDIA	237
Db	188	TVSCLEQGNNSKRTVASTAMNQSSESHAFVISIQRYK-----NDKNSFSRKLHLVDIA	243
Qy	238	GSERAAQTGAAGVRLKEGCNINRSFIILGQVKKLSGQVGGFINRYRDSKLTIRLQNSIG	297
Db	244	GSERQKTKAEGDRLREGINRNGLLCIGNVISALGDGKGNFVPRYRDSKLTIRLQDSLG	303
Qy	298	GNPKTRITCTTPV--SDEFTLTALQFASHTAKYMKNTPYNVNEVSTDEALLKRYRKEIMDL	355
Db	304	GNSHTLMTACVSPADSNLEFTNIRYADRARKTKNKPIN-----IDPQAAELNHL	355
Qy	356	KQLEEVSL-----ETRAQAMEKDQAQLLEEKDLOKQVONEKLEN	396
Db	356	KQVQOQLQILLQAHGGTLPGDINVEPSENIQSLWEXNQ--SLVBENEKLSRGISEAAGQ	413
Qy	397	LTRML-----VTSSSITLQOELKAKKRRVYTWCLGKINK-----MKNSYADQFNIPNTITTK	449
Db	414	TAQMLERIILTEQANEKMAKLEIRRHAAKVDQLKVLTEBQELKENTEILICNLQOV	473
Qy	450	THKLS-----INLLREIDSVCSS-----ESDVESNTDLTISE-----IEWNP	486
Db	474	LAQLSDEAAACMTATITAGEADTVQVSSPDTSRSSDVFS--TQHALRQAQMSKELIELNK	532
Qy	487	A-----TKLINOENIESELNLRADYD-----NLVDYEQLRTEKEMELKLA- EKNDL	534
Db	533	ALALKEALAKKWTQ--DNQLQPIQFYQDNINKLSEVLSLQREKELVLLEQTAKKDA	590
Qy	535	DEFEALERTKKDOEWQLIHEINLKNLVKHRREVYNQDLENELSSKVLIRREKBDQIKKL	594
Db	591	NQAKLSERRRKLQELE--GQIADLKKGLQ-----EQSKLLKKESTEHTVSKL	637
Qy	595	QEYIDSOKLENIKWDLSYSLESIEDPKMKQTLFDFAETVAL---DAKRESAFURSE-NLE	650
Db	638	NQEIRMMKQORVOL-WRQMKEDAEKFPWQKQK-DKEVTLQKERDRKQYVELKLEIRNFQ	695

QY	651	-----LKEKMLATYKQWENDIOLYQSLAEAKKQVVDLEKELQSAFNEITKLTSLID	705
Db	696	KQSNVLRRTTEAAAAANKKDALQ-----KQKVAEKRTQSRGWSTAAARM---	744
QY	706	GKVPKDLNLELEGGKITDLQKELNEKEVEENALREEVILLSELKSLPSEVERLRKBIOD	765
Db	745	-----KWLGN-----EIVVSTEEAKRHNGLEERKILAQDVAQL-KEKRE	787
QY	766	KSELHIITSEKDLFSEVVHKSRSVGLLEETGKTKDLDLATTQSNVKTQDFQNFKL	825
Db	788	SGENPELKLRRRTFSYDEIHQDS---GABDSIAK-----	819
QY	826	HMFQEKYKVLNEENRMOEIVNLSKEAQFDSLSGALKTSLSYKTQELQEKTRVQER	885
Db	820	-----QIESLELELSAQI-----ADLQK	841
QY	886	LNMEQQLKEQENRDSPLQTVREKTLITEKLOOTLEEVKTLTQEKDDLKQLESQIER	945
Db	842	LLD-----AESDRPKQWESATILBAK-----AIKYLVELVSSKILVSKLESSL	889
QY	946	DQLKSDIHDVNNIDTQOLRNALESK-----HQETINTLSKISEEVSRLNH	997
Db	890	NQSKASCIDVQKMLFEEQNHFATETELKEELVKVEQOHOEKVLYLLSQSQSQMTKQL	949
QY	998	BENTGETKDFQKQWGVGIDKKQDLKAKNTOTLTADVKNELIEOQRKIFSLIQEKNELOQ	1057
Db	950	EESVSEK-----EQLLSTLCKQEEELRMQ-----EVCEQNOQ---LLQENSAIKQ	993
QY	1058	MLESV-TAEKELQKTLKENTEMTIENQ-----BELRLGLDELKQKQ	1098
Db	994	KLTLQVASKQ---KPHLTRNIFQSPDSFVIPPKPCKRIKCEQSFAGVGLQYSE	1051
QY	1099	EIVAQEKNAHAKGSELRTCDRLAEVKEKLE-----KSQQLQEKQQLLNQOEE	1149
Db	1052	PSVAEQDNEDGDHAEDEWIPTKLVKSKSIQGCCKGWCNQCQCGCKQKSDCNVCS	1111
QY	1150	MSEQKKNINEENLKNELKNELTLEHETERLELAQKLNENY-----EVKS-----	1197
Db	1112	CDPTK-----CNRHONQD-----NSDAIELNQDSNSFKLEDPTVTSGLSFPH	1157
QY	1198	--ITKERVKLKEL-----QKSFETERDLRGYIREIATGLQTKKEELKIAHIHKE	1246
Db	1158	ICATPSSKILKEMCADQVQLQPMFVSSDH-----PELKSTASESQE	1201
QY	1247	HQETIDELRRSVSEKTAQIINTQDL-BKSH	1275
Db	1202	NKAIGKKKKRALASNTSFSGSGSPIQESHS	1231

RESULT 13
T18372
repeat organellar protein - Plasmodium chabaudi
C:Species: Plasmodium chabaudi
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18372
R:Werner, E.B.; Taylor, W.R.; Holder, A.A.
Mol. Biochem. Parasitol. 94, 185-196, 1998
A:Title: A Plasmodium chabaudi protein contains a repetitive region with a predicted spe
A:Reference number: Z18922; MUID:98418765; PMID:9747969
A:Accession: T18372
A:Status: Preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1939 <WER>
A:Cross-references: EMBL:U43145; NID:g1151157; PID:g1151158; PIDN:AAC63403.1

Query Match 7.1%; Score 946; DB 2; Length 1939;
Best Local Similarity 22.1%; Pred. No. 4.2e-18;
Matches 492; Conservative 438; Mismatches 768; Indels 532; Gaps 105;

QY 542 RYTKKQO-----EMQLIHEISNLKVLKRVNVDLELSSKVELLRKEQIKKLQE 596
Db 6 KSKKNBDGSKDSSKKTNETSGIEKESKNKYNKVNNS-----TKDKDKNDS 56

QY	1597	RVOEALQIBRDQIKENTZEIVAKMKESEKBYQFLK--MTAVNETQEK-----MCEIE	1647
Db	999	GIEEYKVAIBALABEKDVVTKLGEQHQBIATKLEDGHEKVVNEVEKKNASLJMLLEN	1058
QY	1648	HLKEQFETOKNLNENIETIRLQIILHENLEEMRSVTKERDDDLRSVEETLKVERDOLKE	1707
Db	1059	HKNEMI--KLKEEHEKAS-DLVEKLYQKDERVKNSNNKIEETLNVYIKOLNDSIMCYKK	1114
QY	1708	NURETITRLEKQEE--LKIIVHMLHEQHTIDKLRIGVSEKTEINSNOKOLEHSNDA	1764
Db	1115	QILIEVEKRNVEYBEINKLIVQNEKMDN--DKK--IIEKENEIKLKNKLSN-----	1164
QY	1765	LKAQDLKIQEELRIAHMLKEQOETIDKLRIGVSEKTDKLNSMOKOLENSNAKLOEKIOE	1824
Db	1165	YKVFEFK-ENTYKNSMVVNENKER-----IIVDSVCKENISBESDVEGGGNLKWTL-S	1216
QY	1825	LKANEHQILITILKQVNETOKV-----SEMQLKKQIKDOSLTLKSLEINLMAOE	1876
Db	1217	LKKKERNITFSINDKNSESELVDITKSAYNKIBMYKEITEDGNKJIEDJKNKILDSNE	1276
QY	1877	LHENLEEMKSVYKERNDLRRVEETLKLERDQLESQETKARDLEIQOELKTARMLSKEH	1936
Db	1277	L-INLENNKNVLTENN-----NLK-----KEIEKD-----	1302
QY	1937	KETVDKLEKISSEKTIQISDIQDKDLKSDBLQKLOEKLOKBLQLLRVKEDVNMSHKKI	1996
Db	1303	-----NKUNSEKKNENTILANDIIK-----LKEISEMWDREKL--TKENIKLK----	1348
QY	1997	NEMEQLKQFPNVLKCEMDNFOLTCKLHSELEERIVAKERDELRRK-ESLKWEROQ	2055
Db	1349	NDIRQINKEY-----KIKENLMI--KFNENTNEVTSU--KNQIEIEKMKJELKKNVBL	1399
QY	2056	FIATLRE--MIARDQNHQVPEKRLISD--QOQHLMESLREKCSRIKELLKRYSEMDD	2110
Db	1400	LLAEKRETNMSISDNDKNIV--ENNILEDDTSQNNLNKNVEDKTG-----D	1444
QY	2111	HYECLNELSLDLKEIEFHRIMKKLVLSYVTIKBOHECINKPEMDFIDEBKQKEL	2170
Db	1445	DINC--EKNNDQAKESY-----LKDE-----IKKISMLYGEELNRKNSY	1482
QY	2171	LIKIQHLQDCDVPFSRELRDLKNQNDMLHIEILKDFSESEFPSSIKTFEQOVLNRRKM	2230
Db	1483	DEKVQNL-----TNELKELKTR-----NKGE	1504
QY	2231	TOFLEEWLNTFRDIEKLKNGIQKENDRICQVNNFFNNRIIAIMNBST-EFEERSATISKE	2289
Db	1505	EATAE-----LNKLKN-IKEKNKSVAQNDESSNNIITKQDKTPYVSNDQKIQKD	1555
QY	2290	WEODL-KSLKEKNEKLFKNVQTLKTSLASGAQVNPPTQDNKPHVTSRATQITTEKI---	2345
Db	1556	WKANLVLKKEKPD-LWDNINSLE--KENFRVMSIVKENK-----VQNDKIVGI	1602
QY	2346	-----RELENS-----LHEAKESAMHKEKSLIKMKQKELEVNDI--	2379
Db	1603	YSYFKKCEKELKNDMLVICLVLKDILSLFLNDNFVNLFEKIDKI--LWKQMYIPTETIRI	1660
QY	2380	-----TAKLOAKVHESNKCLEKTEKTIQVLODKVALGAKPYKEETEDLKMVLKVIDL	2431
Db	1661	LFURYFSFLDKLRNYV--KCNBEYVNNERYEYSWAL-FQTYLETASNLKKEMYIVVL	1715
QY	2432	EKKMK-----NAKEPEKISATKAT--VEYQKEVIRL-----LRENLRSSQAA-----	2473
Db	1716	EKAOKDSCENNSNFOK-----PKITDILNFSKDSIRLKTIAQLAKELNFRFEAKNILNYD	1771
QY	2474	TSVISEHTDPQPSNKLPTCGGSGGIVQN--TKALILK--SEHIREKEISKLKQOENL-	2528
Db	1772	YQII-----LNKYHECLRLKIVQNMARELDNFYVNSKFSIKKLEMSCDSEDEFK	1823
QY	2529	---IK--QKNELLSNNQH-----LSNEVKTWKER-TLKREAHKQVTCENSPKSPK	2572
Db	1824	YNNIKNNEEKNDITDKPKGNLNIQKIINLQRNKTEKKKNLNVNEINTMYDGTTPKG-K	1982
QY	2573	VTGTASKKKQ	2582

Db 1883 IFTTNDNSKQ 1892

RESULT 14

I51617
kinesin-like protein 1 - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 13-Sep-1996 #sequence revision 13-Sep-1996 #text_change 02-Feb-2001
C:Accession: I51617; A48835; S48837
R:Vernos, I.; Raats, J.; Hirano, T.; Heasman, J.; Karsenti, E.; Wylie, C.
Cell 81, 117-127, 1995
A:Title: Xklp1, a chromosomal Xenopus kinesin-like protein essential for spindle organization
A:Reference number: A56221; MUID:95236444; PMID:7720067
A:Accession: I51617
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1226 <VE>
A:Cross-references: EMBL:X82012; NID:G562792; PIDN:CAA57539.1; PID:G562793
R:Vernos, I.; Heasman, J.; Wylie, C.
Dev. Biol. 157, 232-239, 1993
A:Title: Multiple kinesin-like transcripts in Xenopus oocytes.
A:Reference number: A48835; MUID:93246065; PMID:8482413
A:Accession: A48835
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 9-162, 'L', 164-338 <VE2>
A:Experimental source: oocyte
A:Note: sequence extracted from NCBI backbone (NCBIP:130975)
C:Genetics:
A:Gene: klp1
C:Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology
C:Keywords: ATP; nucleotide binding; P-loop
F:9-343/Domain: kinesin motor domain homology <KMOT>
F:187-94/Region: nucleotide-binding motif A (P-loop)

Query Match 7.0%; Score 935.5; DB 2; Length 1226;
Best Local Similarity 24.6%; Pred. No. 5.1e-18;
Matches 341; Conservative 261; Mismatches 478; Indels 309; Gaps 45;

Qy	3	ERG-AVAVCVRRVRPLNSREESLG-ETAQYMKTDNNVYQVDSKSFNDFRVFHGNETTK	60
Db	4	DEGIPVRVALRCRPLVPKENNEGCKMCLTFVPGEQQVI--VGTEKSFTYDYVDFSAEQE	61
Qy	61	NYIEEIAAPIIDSALQGYNGTIFAYGQTASGKTYTMGSEDH-----LGVIPRAIHD	113
Db	62	EVYNSAVAPLKGIFPKGNVAVLAYQGTGSGKTYSMWGAYTHQNEPFTVGVPKPTVIAL	121
Qy	114	FQKIKKFPDRFLRVSTMYBINEYETITDLL-CGTQMKPLIIREDVNRNVYVADLITEVV	172
Db	122	FREIHQRPWEFNLKVSYLEYINBEILDLLVAARDKNTNISREDPKGKICGLTERDV	181
Qy	173	YTSWALKWIPKGEKSRHYGTCKNQRSSRSHITFRMTLESREKCEPNCESGVKVSJLN	232
Db	182	KTALDITLSCLEQNSRTRVASTAMNSQSSRSHAFITISIEQRKEDGKN---SFR-SKLH	237
Qy	233	LVDLAGSERAAQTGAAGVRLKEGNCINRSLFILGOVIKGLSD-GOVGGFINVRDSKLTRI	291
Db	238	LVDLAGSERQKTKRAEGDRLKEGISINRGLLCLGNVISALGDESKGGFVPRDSKLTRL	297
Qy	292	LONSLGNPKTRIICITIPV--SFDETITALQFASKYMKNTVPYVNEVSTDEALLKRYR	349
Db	298	LQDSLGNSHITLMTACVSPADSNMEETLNTLRYADRARKIKNKPVN-TDPAQAELORLK	356
Qy	350	KEIMDLKKQLEEV-----SLETRAQAMEKD-----LQAQ	378
Db	357	LOVOELQVLLQAHGGTLPVLSNMEPSENLQSLMERNKNLEKENGKLSRELGEAAVQTAQ	416
Qy	379	LLEKEDLLQKVQNEKI-----ENLTRMLVTSSSLTLO-----QELKA	415
Db	417	FL-EKIIMTQQQNEKGSKMEELKQHAACKVNLQRLVETLEQBELKONVEIQNLQOVIV	475
Qy	416	KRKRRVTWCLGKNKMNKSNYAQFNIP-----TWITTKTKLSINLREI	461

```

476 QLODESSGAGSIAMDEB--AASFVPPEDESGEKSSDGFTHHALROAQSKELI-EL 532
462 DESVCSSEDFNTLTLSEIWNPAKLLNQNTESELSRADYDNLVLDYEQRLTEK 521
533 NKALVMKEALAKMAQNDQLEPIOSEYLNKHLSEVGLQKEBELILALHSAKON 592
522 EEWELKLEKNDLDEPEALERTKQDQEMQLIHEISNLKLVKGRE-----VYNQDLE 574
593 NOAKLSERRKRRLQELGOMTELK-----KGEQSKLLKLRSTKTVAKNQEIQ 644
575 NELSSSVELLREKEDQIKLOEYIDSKLENIMKMDLSYLESIEDPKMQKTLFDATVA 634
645 GMKQORVQLMRQMKDAEFRTWKQKTEVIQL-----KEDKRQVLEI----- 690
635 LDAKRESAFIRSENLEKEMKELATYKQENDIQLYQSOLAEAKKMQVDLEKELQSAF 694
691 ---KLERDFQKQANV-LRRKTEEAASANKLQKALQKQKAMERKQDSQ---SKMGEAA 743
695 NEITKLTSLDGKVPKOLLN-LELEGKITDLQKELNKEVEEALREEVILLSELKSLP 753
744 SRV-----KNWLANEVEVLSTEEAQRHLNLDLEDRKILAQDI---AQLKQKT 788
754 SEYERLKEIQDKSEELHIITSEKDLFSEVVKESRVQGLLEIGIKTKDLDLQTSNYK 813
789 DAGERIPTKIRRT-----YTVAELENLEBEASVTK----- 819
814 STDQEFONFKTLHMDPEQKVMVLEBNERMNOELVNLSEAKQFDDSLGALKTELSYKQ 873
820 -----QIESLETMEELRSA 833
874 ELOKTEVROERLNEBOLKEQLENRDSPLQTVEREKTLITEKLOQTLERVK---TLTQE 930
834 QIADLOQLKLDAGEEEMVWRWETISNIMEAKALKYLITELVSSKVGSKLESSVKQN 893
931 KDLKQIQESLQTERDQKSDIHTVNMNIDTQELRNALLESKQHQETINTLKSISE- 989
894 RAHVADLQKNIFEERNQW-----AEMTEHQSQL---MQLEQHQBKILYLLSLOQK 943
990 EVSNLMEWENTGETKDFQKQMV-----GTDKKQDLAEKNTQTLTADVKDNEIIEQ 1041
944 QASVPVTEELPABEITEREKQMLERLKFQDEIEKMKALCEKQKQLL-----QENDMYKQ 999
1042 QKTFSLIQEKNELQOMLESVIAEKEQLKTD-----LKENIEMTIENQBELR 1088
1000 KLALLHVASGK-----LHNILPRAEICSPPDFIPPKRGKRTNKAASAVILEDL- 1053
1089 LLGDELKQOBIVAOEKXNHAIKKESGELSRDRLAEVEEKL---KEKSOLOEKQOQLLV 1146
1054 LSESEBESDDKNWEPGNNSKQKLTSCSKCARCGNKNMGCRKTKQ-----NC 1104
1147 QEMSEMKKINEIENLKNELKNKELTLEHMETELELAQKLNENYEVKS----- 1197
1105 SDDCFDPSKCRNRDNHDEGHEDQSLE-SENSKID-----YPDVTAGGFPTPC 1155
1198 ITKERVKLKELQKSPETERDHLRGVIRIEATGLQTKBELKIAHILKHEQETIDELRRS 1257
1156 VTPYKVLRLISD-----IGQVLSIKLQKSPSTASASVMSQEN-----QTS 1199
1258 VSEKTAQII 1266
1200 ILTKKKKVL 1208

```

RESULT 15
 T49451
 N:Alternate names: protein B14D6.30
 C:Species: Neurospora crassa
 C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
 C:Accession: T49451
 R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
 submitted to the Protein Sequence Database, May 2000

```

A:Reference number: Z25022
A:Accession: T49451
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1742 <SCH>
A:Cross-references: EMBL:AL356173; GSPDB:GN00116; NCSP:B14D6.30
A:Experimental source: BAC clone B14D6; strain OR74A
C:Genetics:
A:Gene: NCSP:B14D6.30
A:Map position: 6
A:Introns: 58/1; 166/1; 267/3; 1543/3

Query Match      7.0%; Score 929.5; DB 2; Length 1742;
Best Local Similarity 21.5%; Pred. No. 1e-17;
Matches 452; Conservative 351; Mismatches 683; Indels 621; Gaps 81;

6 AVAVCVVRVP-----LNSRESLGETAQVYWKTDNNVYQVDSGSFNF 49
54 AVRVAVRVPPLGPDGPDGVDLPQRFQSRMVQVQGETGV-----AIDSPQGRKLFVF 105
50 DRVFHGNETTKVYVEIAPIIDSAIQVNGTIFAYGOTASGKTYM--MGSE-----DH 102
106 DRVF-GPEVDQEGWBYLSDCVNAFTQGYNVSLLAQYQSGAGKSYTGTAGPDVQEDLEA 164
103 LGVIPRAIHDIPOKIK-----KFP-----DREFLLRVSYM 132
165 MGVIPRAALFELKDGSSPKSQGAASKSSMSQLRAPSRVAMLPQSNIDKDWKLATATYV 224
133 EYINETITDLC-----GTQMKPLIIRDVNRNVVADLTVEVYVTSSEALKWIT 183
225 EYINETLRLIPEHIPQHERGT-----VTIRBDVKNIIITGLQ-----V 266
184 KGKSRHYGETKYNQSSRSSTHTIFRMLESREKGP-----SNCEGSVKV 228
267 EGSALRQTDATANAKSSSSSHAVFSNLVRKAGKPTAPTDRRMSMLEAMSGTEAMVTT 326
229 -SHNLVLDLAGSRAAQTGAAGVRLKEGCNINRSLFILGVQIKLSQGVGGFNYNRDSK 287
327 DSMHFDVLDLAGSELKNTGQGERAKSGISINAGLAALGKVISQLSSRPGCAHVSYRDSK 386
288 LTRILQNSLGNPKTRIICITIPVSF--DETALQFASAKYMKNTFYNEV--STDEA 343
387 LTRLLQDSLGNATYTIACVTQAEPLHSETLNTVQVQARARATQKPERIQQVDEGKQA 446
344 LLKRYKIEIMDKKQLESEVSLTEAQAMEKDQALAEKDLQKQVNEKINENITRMVLT 403
447 IIERLKAQVAFLEQIRS-----SERGGDRRRNLLAPGERSER-QNERAEALQNLQD 499
404 S--SSLTLQBELKAKRRRTVTCIGKINKMKNYADQFNIPNTITTKHKLINLLREI 461
500 ARENYTILSQ-----RHAKLISEMAKARENEFAEN-----QHL 532
462 DESVCSSEDFNTLDTLSEIWNPAKLLNQNTESELSRADYDNLVLDYEQRLTEK 521
533 EESL-GES-----ATERLNRS-----NSFAQAVEQVWLEYE----- 562
522 EEMELKLEKNDLDEFEALERKTKQDQEMQLIHEISNLKLVKREVYNQDLENELSKV 581
563 -----KT-----IQSLEQSLAS-- 574
582 ELLREKEBQIKKLEBYIDSQKLENIMKMDLSYLESIEDPKMQKTLFDATVALDAKRES 641
575 -----TRATLANTTEATLLEKETKC 593
642 AFLRSNLELKEKMKELATTYKQENDIQLYQSOLAEAKKMQVDLEKELQSAFNITKLT 701
594 AYTETINTQQLARQKLMREASTEN---YLHDLER-----KLD 629
702 SLIDGKVPKOLLNLEBKITDLQKELNKEVEEALREEVILLSELKSLPSEVERLRK 761
630 SHTSGEERKNAT-----ITELREKFIAR-VRENEANAEDYISTLE-----ERLAE 672
762 EIQDKSEELHIITSEKDLFSEVVKH-----SRVQGLLEIGIKTKDLDLQTSNYKSTQDE 818

```

Db 673 A00D-----AELM0REIDRL-EQVIERQSRIGKLDLSLNLDELHIQ0DPPT-----716
QY 819 FQNFKTLHMDFEQYKMWLEENRMQEIWN---LSKEAQKFDSSIGALKTSLSYKTQEL 875
Db 717 -----IPENH---HEUTNGRHKMASRDYANHSRQSHVSHRSQ-F 754
QY 876 QKTRVQRLNMEQKQLE-NRD-SPLQTVREKTLITELKLOOT-----LERVKTL 927
Db 755 DEPIRESSE-----EDIPEDENLDATPTDKANTPKASHAQLAGKNGSKLVPDQFETV 809
QY 928 TOEKDILKQLOSLQIERQKSDIDHTVMNIDTOEQLRNALLESKQHQETINTLK---984
Db 810 TKELVDL-----RTEHETTLHDYSGALQAKHEE---ALRALAEQDAVDEARHPS 855
QY 985 ----SKISEVSRNLHMEENTGETDEFOQKMGVGDKKQDLKAKTQTLTADV-----1034
Db 856 RYDLSILSVSAGTRPMSLMSGETKGRSLSELS-----SALGSHTVVTDVSGAETAK 910
QY 1035 --DNEIIBQORKIFSLIQEKNELQOMLESVIAEKQLKTDLKENIEMTIENOEELRLIGD 1092
Db 911 HHDSEIDE-----ADADAQIQKLFIAAKAABERELATRYAQLKEKHQETLDMVE 961
QY 1093 ELKQOEIVAOEKHAIKEGELSRTCDRLAEVEKLEKESQO-----LQE 1138
Db 962 ELKTE---IA-KAQLSVSESSIRSTPV-----IRKSSQNVMIIDRAHRSFATLSN 1010
QY 1139 KOQQLNVQEESEMOKKINETENLKNELKNELTLEHMETERLELAOKLKNENYEVKSI 1198
Db 1011 IAAENLHDPDVWAFELNLA---AMHHLHARSVRIQELAEADVNAKXEMETKMTIISGL 1067
QY 1199 TKERKVLK-----ELQKSPETERDLRGVIREI-EATGLQTK-E-LKIAHILK 1245
Db 1068 TRERSLLKAAAFMDMAMVSSLRQELERSEKH---VQELRESNAIRQRELEAQISELOHA 1123
QY 1246 EHOETID-----ELRSVSEKTAQIINTQDLEKSHTKLOE 1280
Db 1124 LHQAGVDPASVGPVDFGDSDAQAKRIVELEAEALRGWEAKHQAALENLQD---SENRMKS 1180
QY 1281 EIPVLHBEQELLPNVKVSETQETMNELELLTEQSTTKDSTTLARIEMERLRINEKFOES 1340
Db 1181 SMAELGSIQSLFTTQLETETQSSML-----SSRLAQTEGQVTSINDQLAQN 1227
QY 1341 QBEIKSLTKERDNLTKIKEALEVHDQDKHEHIRETLAKIQESQSQKQESQSLNMKEKDNETT 1400
Db 1228 GH-----KEPEGDEAAKEK---KHKELVEVLRREIDDYKEIVNRTQTKVTQTEQAAAT 1278
QY 1401 KIVSEMEQFKPKDSALLRIETEMGLSKLOESHDDEKMSVAK-----EKDDQLRLQ 1451
Db 1279 KVLIDAT-RERDDAAABAE-HNKQLVNRLEESITEHEHTIKAHQESLHALELNALELE 1336
QY 1452 EVLQESDQKAKENIKIVAKHLETEELKVAHCCLKEQEEITINELRVNLSEKETEISTIQ 1511
Db 1337 EARTASRREVEAEMKAKAHAEQVRL-----EGLTDAREDLLRVATQVA-LA 1385
QY 1512 KQLBAINDKQNKTOEIVYKEEQLNI---KQISVQENVNVLKOPKEHKKAKMSALQSIES 1569
Db 1386 LGLEVSTELKSLERIEDLLESRDQVDLEQKRGLEQHVVELTSI-----NDQVWNELES 1439
QY 1570 KMLLETNRLQESQBEIQIMIKEEMKRVQPALQIERDQLKENTKEIVAKMKESQEKYQ 1629
Db 1440 VKAALADMLATENDKVKTVVPVKESVALVK-----KMWDLTK---1478
QY 1630 FLKMTAVNETQCKMCELEHLEKEQFETQKLNLENIETENIRLTOILHENLEEMRSVTYKRD 1699
Db 1479 -----NKNNSRL---VEELEDQLOS---NYDQVQITSNRLSMLOSEKTOOLEEANA--1524
QY 1690 DLRSVEETLKVRED-----QLKENLRETIITRDLEKQEBELKIVH-----MHLKE 1732
Db 1525 -IIRLEELKWARREYAAALQTKYNVLSQEAASVPQRSKSPAPNPSPIRKSNSVQSLPS 1583
QY 1733 HQETI-----DKLRGIVSEKTNNEISNMOKDLEHSNDALKQAQDLKTOEELRIAHMHLKE 1785

Db 1584 PPPAIPPLPLPSDRASGAVSPGAPRPSKDILAA--ALSNOQFOAIOEDQEARIRIE 1641
QY 1786 QOETIDKLRGIVSEKTDKLSNMQKOLENSNAKLOEKIQBELKANEHQLITLKKDVNETQKK 1845
Db 1642 -----KNLLAEK-----QLTATLLEALITDL---ETQOLKIKADADAKRR 1678
QY 1846 VSEMEQKQKIQKQOSITLSKLEIENLNLAQELHENLEEMKSVMKERDNLRRVEETLKLER 1905
Db 1679 AGELEABLKELKD-----KPOVDNRWSLQQVEEERKKRVDAERARQHLLEERMQSLASGK 1732
QY 1906 DQKESL 1912
Db 1733 KKKKGS 1739

Search completed: July 29, 2004, 09:41:00
Job time : 61.7887 secs

Blank sheet
USPTO